



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C12N 15/31, C07K 14/22, 16/12, A61K 39/095 // (C12N 15/31, C12R 1:36)</b>		<b>A2</b>	(11) International Publication Number: <b>WO 99/24578</b>
			(43) International Publication Date: 20 May 1999 (20.05.99)
(21) International Application Number: <b>PCT/IB98/01665</b>		(74) Agent: <b>HALLYBONE, Huw, George; Carpmaels &amp; Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).</b>	
(22) International Filing Date: <b>9 October 1998 (09.10.98)</b>			
(30) Priority Data:		(81) Designated States: <b>AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</b>	
9723516.2      6 November 1997 (06.11.97)      GB 9724190.5      14 November 1997 (14.11.97)      GB 9724386.9      18 November 1997 (18.11.97)      GB 9725158.1      27 November 1997 (27.11.97)      GB 9726147.3      10 December 1997 (10.12.97)      GB 9800759.4      14 January 1998 (14.01.98)      GB 9819016.8      1 September 1998 (01.09.98)      GB			
(71) Applicant (for all designated States except US): <b>CHIRON S.P.A. [IT/IT]; Via Fiorentina, 1, I-53100 Siena (IT).</b>			
(72) Inventors; and		<b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>	
(75) Inventors/Applicants (for US only): <b>MASIGNANI, Vega [IT/IT]; Via Pantaneto, 105, I-53100 Siena (IT). RAP-PUOLI, Rino [IT/IT]; Via delle Rocche, 1, Vagliagli, I-53019 Castelnuovo Berardenga (IT). PIZZA, Mariagrazia [IT/IT]; Strada di Montalbuccio, 160, I-53100 Siena (IT). SCARLATO, Vincenzo [IT/IT]; Via Firenze, 3/37, I-53134 Colle Val d'Elsa (IT). GRANDI, Guido [IT/IT]; 9° Strada, 4, I-20090 Segrate (IT).</b>			
(54) Title: <b>NEISSERIAL ANTIGENS</b>			
(57) Abstract			
<p>The invention provides proteins from <i>Neisseria meningitidis</i> (strains A and B) and from <i>Neisseria gonorrhoeae</i> including amino acid sequences, the corresponding nucleotide sequences, expression data, and serological data. The proteins are useful antigens for vaccines, immunogenic compositions, and/or diagnostics.</p>			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						



## NEISSERIAL ANTIGENS

This invention relates to antigens from *Neisseria* bacteria.

### BACKGROUND ART

5 *Neisseria meningitidis* and *Neisseria gonorrhoeae* are non-motile, gram negative diplococci that are pathogenic in humans. *N.meningitidis* colonises the pharynx and causes meningitis (and, occasionally, septicaemia in the absence of meningitis); *N.gonorrhoeae* colonises the genital tract and causes gonorrhea. Although colonising different areas of the body and causing completely different diseases, the two pathogens are closely related, although one feature that clearly differentiates meningococcus from gonococcus is the presence of a polysaccharide capsule that is  
10 present in all pathogenic meningococci.

*N.gonorrhoeae* caused approximately 800,000 cases per year during the period 1983-1990 in the United States alone (chapter by Meitzner & Cohen, "Vaccines Against Gonococcal Infection", In: *New Generation Vaccines*, 2nd edition, ed. Levine, Woodrow, Kaper, & Cobon, Marcel Dekker, New York, 1997, pp.817-842). The disease causes significant morbidity but limited mortality.  
15 Vaccination against *N.gonorrhoeae* would be highly desirable, but repeated attempts have failed. The main candidate antigens for this vaccine are surface-exposed proteins such as pili, porins, opacity-associated proteins (Opas) and other surface-exposed proteins such as the Lip, Laz, IgA1 protease and transferrin-binding proteins. The lipooligosaccharide (LOS) has also been suggested as vaccine (Meitzner & Cohen, *supra*).

20 *N.meningitidis* causes both endemic and epidemic disease. In the United States the attack rate is 0.6-1 per 100,000 persons per year, and it can be much greater during outbreaks (see Lieberman *et al.* (1996) Safety and Immunogenicity of a Serogroups A/C *Neisseria meningitidis* Oligosaccharide-Protein Conjugate Vaccine in Young Children. *JAMA* 275(19):1499-1503; Schuchat *et al* (1997) Bacterial Meningitis in the United States in 1995. *N Engl J Med* 337(14):970-  
25 976). In developing countries, endemic disease rates are much higher and during epidemics incidence rates can reach 500 cases per 100,000 persons per year. Mortality is extremely high, at 10-20% in the United States, and much higher in developing countries. Following the introduction of the conjugate vaccine against *Haemophilus influenzae*, *N. meningitidis* is the major cause of bacterial meningitis at all ages in the United States (Schuchat *et al* (1997) *supra*).

Based on the organism's capsular polysaccharide, 12 serogroups of *N.meningitidis* have been identified. Group A is the pathogen most often implicated in epidemic disease in sub-Saharan Africa. Serogroups B and C are responsible for the vast majority of cases in the United States and in most developed countries. Serogroups W135 and Y are responsible for the rest of the cases in the United States and developed countries. The meningococcal vaccine currently in use is a tetravalent polysaccharide vaccine composed of serogroups A, C, Y and W135. Although efficacious in adolescents and adults, it induces a poor immune response and short duration of protection, and cannot be used in infants [eg. Morbidity and Mortality weekly report, Vol.46, No. RR-5 (1997)]. This is because polysaccharides are T-cell independent antigens that induce a weak immune response that cannot be boosted by repeated immunization. Following the success of the vaccination against *H.influenzae*, conjugate vaccines against serogroups A and C have been developed and are at the final stage of clinical testing (Zollinger WD "New and Improved Vaccines Against Meningococcal Disease" in: *New Generation Vaccines, supra*, pp. 469-488; Lieberman *et al* (1996) *supra*; Costantino *et al* (1992) Development and phase I clinical testing of a conjugate vaccine against meningococcus A and C. *Vaccine* 10:691-698).

Meningococcus B remains a problem, however. This serotype currently is responsible for approximately 50% of total meningitis in the United States, Europe, and South America. The polysaccharide approach cannot be used because the menB capsular polysaccharide is a polymer of  $\alpha(2-8)$ -linked *N*-acetyl neuraminic acid that is also present in mammalian tissue. This results in tolerance to the antigen; indeed, if an immune response were elicited, it would be anti-self, and therefore undesirable. In order to avoid induction of autoimmunity and to induce a protective immune response, the capsular polysaccharide has, for instance, been chemically modified substituting the *N*-acetyl groups with *N*-propionyl groups, leaving the specific antigenicity unaltered (Romero & Outschoorn (1994) Current status of Meningococcal group B vaccine candidates: capsular or non-capsular? *Clin Microbiol Rev* 7(4):559-575).

Alternative approaches to menB vaccines have used complex mixtures of outer membrane proteins (OMPs), containing either the OMPs alone, or OMPs enriched in porins, or deleted of the class 4 OMPs that are believed to induce antibodies that block bactericidal activity. This approach produces vaccines that are not well characterized. They are able to protect against the homologous strain, but are not effective at large where there are many antigenic variants of the outer membrane proteins. To overcome the antigenic variability, multivalent vaccines containing up to nine different

porins have been constructed (eg. Poolman JT (1992) Development of a meningococcal vaccine. *Infect. Agents Dis.* 4:13-28). Additional proteins to be used in outer membrane vaccines have been the opa and opc proteins, but none of these approaches have been able to overcome the antigenic variability (eg. Ala'Aldeen & Borriello (1996) The meningococcal transferrin-binding proteins 1 and 2 are both surface exposed and generate bactericidal antibodies capable of killing homologous and heterologous strains. *Vaccine* 14(1):49-53).

A certain amount of sequence data is available for meningococcal and gonococcal genes and proteins (eg. EP-A-0467714, WO96/29412), but this is by no means complete. The provision of further sequences could provide an opportunity to identify secreted or surface-exposed proteins that are presumed targets for the immune system and which are not antigenically variable. For instance, some of the identified proteins could be components of efficacious vaccines against meningococcus B, some could be components of vaccines against all meningococcal serotypes, and others could be components of vaccines against all pathogenic *Neisseriae*.

## THE INVENTION

The invention provides proteins comprising the Neisserial amino acid sequences disclosed in the examples. These sequences relate to *N.meningitidis* or *N.gonorrhoeae*.

It also provides proteins comprising sequences homologous (*ie.* having sequence identity) to the Neisserial amino acid sequences disclosed in the examples. Depending on the particular sequence, the degree of identity is preferably greater than 50% (eg. 65%, 80%, 90%, or more). These homologous proteins include mutants and allelic variants of the sequences disclosed in the examples. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between the proteins is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affine gap search with parameters *gap open penalty=12* and *gap extension penalty=1*.

The invention further provides proteins comprising fragments of the Neisserial amino acid sequences disclosed in the examples. The fragments should comprise at least *n* consecutive amino acids from the sequences and, depending on the particular sequence, *n* is 7 or more (eg. 8, 10, 12, 14, 16, 18, 20 or more). Preferably the fragments comprise an epitope from the sequence.

The proteins of the invention can, of course, be prepared by various means (*eg.* recombinant expression, purification from cell culture, chemical synthesis *etc.*) and in various forms (*eg.* native, fusions *etc.*). They are preferably prepared in substantially pure or isolated form (*ie.* substantially free from other Neisserial or host cell proteins)

- 5 According to a further aspect, the invention provides antibodies which bind to these proteins. These may be polyclonal or monoclonal and may be produced by any suitable means.

According to a further aspect, the invention provides nucleic acid comprising the Neisserial nucleotide sequences disclosed in the examples. In addition, the invention provides nucleic acid comprising sequences homologous (*ie.* having sequence identity) to the Neisserial nucleotide  
10 sequences disclosed in the examples.

Furthermore, the invention provides nucleic acid which can hybridise to the Neisserial nucleic acid disclosed in the examples, preferably under "high stringency" conditions (*eg.* 65°C in a 0.1xSSC, 0.5% SDS solution).

Nucleic acid comprising fragments of these sequences are also provided. These should comprise  
15 at least  $n$  consecutive nucleotides from the Neisserial sequences and, depending on the particular sequence,  $n$  is 10 or more (*eg.* 12, 14, 15, 18, 20, 25, 30, 35, 40 or more).

According to a further aspect, the invention provides nucleic acid encoding the proteins and protein fragments of the invention.

It should also be appreciated that the invention provides nucleic acid comprising sequences  
20 complementary to those described above (*eg.* for antisense or probing purposes).

Nucleic acid according to the invention can, of course, be prepared in many ways (*eg.* by chemical synthesis, from genomic or cDNA libraries, from the organism itself *etc.*) and can take various forms (*eg.* single stranded, double stranded, vectors, probes *etc.*).

In addition, the term "nucleic acid" includes DNA and RNA, and also their analogues, such as  
25 those containing modified backbones, and also peptide nucleic acids (PNA) *etc.*

According to a further aspect, the invention provides vectors comprising nucleotide sequences of the invention (eg. expression vectors) and host cells transformed with such vectors.

According to a further aspect, the invention provides compositions comprising protein, antibody, and/or nucleic acid according to the invention. These compositions may be suitable as vaccines, for instance, or as diagnostic reagents, or as immunogenic compositions.

The invention also provides nucleic acid, protein, or antibody according to the invention for use as medicaments (eg. as vaccines) or as diagnostic reagents. It also provides the use of nucleic acid, protein, or antibody according to the invention in the manufacture of: (i) a medicament for treating or preventing infection due to Neisserial bacteria; (ii) a diagnostic reagent for detecting the presence of Neisserial bacteria or of antibodies raised against Neisserial bacteria; and/or (iii) a reagent which can raise antibodies against Neisserial bacteria. Said Neisserial bacteria may be any species or strain (such as *N.gonorrhoeae*, or any strain of *N.meningitidis*, such as strain A, strain B or strain C).

The invention also provides a method of treating a patient, comprising administering to the patient a therapeutically effective amount of nucleic acid, protein, and/or antibody according to the invention.

According to further aspects, the invention provides various processes.

A process for producing proteins of the invention is provided, comprising the step of culturing a host cell according to the invention under conditions which induce protein expression.

A process for producing protein or nucleic acid of the invention is provided, wherein the the protein or nucleic acid is synthesised in part or in whole using chemical means.

A process for detecting polynucleotides of the invention is provided, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridizing conditions to form duplexes; and (b) detecting said duplexes.

A process for detecting proteins of the invention is provided, comprising the steps of: (a) contacting an antibody according to the invention with a biological sample under conditions suitable for the formation of an antibody-antigen complexes; and (b) detecting said complexes.

A summary of standard techniques and procedures which may be employed in order to perform the invention (eg. to utilise the disclosed sequences for vaccination or diagnostic purposes) follows. This summary is not a limitation on the invention but, rather, gives examples that may be used, but are not required.

## 5 General

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature eg. Sambrook *Molecular Cloning; A Laboratory Manual, Second Edition* (1989); *DNA Cloning, Volumes I and*  
10 *ii* (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); *Transcription and Translation* (B.D. Hames & S.J. Higgins eds. 1984); *Animal Cell Culture* (R.I. Freshney ed. 1986); *Immobilized Cells and Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide to Molecular Cloning* (1984); the *Methods in Enzymology* series (Academic Press, Inc.), especially volumes 154 & 155; *Gene*  
15 *Transfer Vectors for Mammalian Cells* (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Mayer and Walker, eds. (1987), *Immunochemical Methods in Cell and Molecular Biology* (Academic Press, London); Scopes, (1987) *Protein Purification: Principles and Practice*, Second Edition (Springer-Verlag, N.Y.), and *Handbook of Experimental Immunology, Volumes I-IV* (D.M. Weir and C. C. Blackwell eds 1986).

20 Standard abbreviations for nucleotides and amino acids are used in this specification.

All publications, patents, and patent applications cited herein are incorporated in full by reference. In particular, the contents of UK patent applications 9723516.2, 9724190.5, 9724386.9, 9725158.1, 9726147.3, 9800759.4, and 9819016.8 are incorporated herein.

## Definitions

25 A composition containing X is "substantially free of" Y when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95% or even 99% by weight.

The term "comprising" means "including" as well as "consisting" eg. a composition "comprising" X may consist exclusively of X or may include something additional to X, such as X+Y.

The term "heterologous" refers to two biological components that are not found together in nature. The components may be host cells, genes, or regulatory regions, such as promoters. Although the heterologous components are not found together in nature, they can function together, as when a promoter heterologous to a gene is operably linked to the gene. Another example is where a Neisserial sequence is heterologous to a mouse host cell. A further examples would be two epitopes from the same or different proteins which have been assembled in a single protein in an arrangement not found in nature.

An "origin of replication" is a polynucleotide sequence that initiates and regulates replication of polynucleotides, such as an expression vector. The origin of replication behaves as an autonomous unit of polynucleotide replication within a cell, capable of replication under its own control. An origin of replication may be needed for a vector to replicate in a particular host cell. With certain origins of replication, an expression vector can be reproduced at a high copy number in the presence of the appropriate proteins within the cell. Examples of origins are the autonomously replicating sequences, which are effective in yeast; and the viral T-antigen, effective in COS-7 cells.

A "mutant" sequence is defined as DNA, RNA or amino acid sequence differing from but having sequence identity with the native or disclosed sequence. Depending on the particular sequence, the degree of sequence identity between the native or disclosed sequence and the mutant sequence is preferably greater than 50% (eg. 60%, 70%, 80%, 90%, 95%, 99% or more, calculated using the Smith-Waterman algorithm as described above). As used herein, an "allelic variant" of a nucleic acid molecule, or region, for which nucleic acid sequence is provided herein is a nucleic acid molecule, or region, that occurs essentially at the same locus in the genome of another or second isolate, and that, due to natural variation caused by, for example, mutation or recombination, has a similar but not identical nucleic acid sequence. A coding region allelic variant typically encodes a protein having similar activity to that of the protein encoded by the gene to which it is being compared. An allelic variant can also comprise an alteration in the 5' or 3' untranslated regions of the gene, such as in regulatory control regions (eg. see US patent 5,753,235).

#### Expression systems

The Neisserial nucleotide sequences can be expressed in a variety of different expression systems; for example those used with mammalian cells, baculoviruses, plants, bacteria, and yeast.

### i. Mammalian Systems

Mammalian expression systems are known in the art. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, usually located 25-30 base pairs (bp) upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element, usually located within 100 to 200 bp upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation [Sambrook et al. (1989) "Expression of Cloned Genes in Mammalian Cells." In *Molecular Cloning: A Laboratory Manual*, 2nd ed.].

Mammalian viral genes are often highly expressed and have a broad host range; therefore sequences encoding mammalian viral genes provide particularly useful promoter sequences. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter (Ad MLP), and herpes simplex virus promoter. In addition, sequences derived from non-viral genes, such as the murine metallothionein gene, also provide useful promoter sequences. Expression may be either constitutive or regulated (inducible), depending on the promoter can be induced with glucocorticoid in hormone-responsive cells.

The presence of an enhancer element (enhancer), combined with the promoter elements described above, will usually increase expression levels. An enhancer is a regulatory DNA sequence that can stimulate transcription up to 1000-fold when linked to homologous or heterologous promoters, with synthesis beginning at the normal RNA start site. Enhancers are also active when they are placed upstream or downstream from the transcription initiation site, in either normal or flipped orientation, or at a distance of more than 1000 nucleotides from the promoter [Maniatis et al. (1987) *Science* 236:1237; Alberts et al. (1989) *Molecular Biology of the Cell*, 2nd ed.]. Enhancer elements derived from viruses may be particularly useful, because they usually have a broader host range. Examples include the SV40 early gene enhancer [Dijkema et al (1985) *EMBO J.* 4:761] and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus [Gorman et al. (1982b) *Proc. Natl. Acad. Sci.* 79:6777] and from human cytomegalovirus [Boshart et al. (1985) *Cell* 41:521]. Additionally, some enhancers are regulatable and become active only



in the presence of an inducer, such as a hormone or metal ion [Sassone-Corsi and Borelli (1986) *Trends Genet.* 2:215; Maniatis et al. (1987) *Science* 236:1237].

A DNA molecule may be expressed intracellularly in mammalian cells. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in mammalian cells. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The adenovirus tripartite leader is an example of a leader sequence that provides for secretion of a foreign protein in mammalian cells.

Usually, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-transcriptional cleavage and polyadenylation [Birnstiel et al. (1985) *Cell* 41:349; Proudfoot and Whitelaw (1988) "Termination and 3' end processing of eukaryotic RNA. In *Transcription and splicing* (ed. B.D. Hames and D.M. Glover); Proudfoot (1989) *Trends Biochem. Sci.* 14:105]. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator/polyadenylation signals include those derived from SV40 [Sambrook et al (1989) "Expression of cloned genes in cultured mammalian cells." In *Molecular Cloning: A Laboratory Manual*].

Usually, the above described components, comprising a promoter, polyadenylation signal, and transcription termination sequence are put together into expression constructs. Enhancers, introns with functional splice donor and acceptor sites, and leader sequences may also be included in an expression construct, if desired. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as mammalian cells or bacteria. Mammalian replication systems include those derived from animal

viruses, which require trans-acting factors to replicate. For example, plasmids containing the replication systems of papovaviruses, such as SV40 [Gluzman (1981) *Cell* 23:175] or polyomavirus, replicate to extremely high copy number in the presence of the appropriate viral T antigen. Additional examples of mammalian replicons include those derived from bovine papillomavirus and Epstein-Barr virus. Additionally, the replicon may have two replicaton systems, thus allowing it to be maintained, for example, in mammalian cells for expression and in a prokaryotic host for cloning and amplification. Examples of such mammalian-bacteria shuttle vectors include pMT2 [Kaufman et al. (1989) *Mol. Cell. Biol.* 9:946] and pHEBO [Shimizu et al. (1986) *Mol. Cell. Biol.* 6:1074].

10 The transformation procedure used depends upon the host to be transformed. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.

15 Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (eg. Hep G2), and a number of other cell lines.

## 20 ii. Baculovirus Systems

The polynucleotide encoding the protein can also be inserted into a suitable insect expression vector, and is operably linked to the control elements within that vector. Vector construction employs techniques which are known in the art. Generally, the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site for insertion of the heterologous gene or genes to be expressed; a wild type baculovirus with a sequence homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media.

25 After inserting the DNA sequence encoding the protein into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques

are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit). These techniques are generally known to those skilled in the art and fully described in Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) (hereinafter "Summers and Smith").

Prior to inserting the DNA sequence encoding the protein into the baculovirus genome, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are usually assembled into an intermediate transplacement construct (transfer vector). This construct may contain a single gene and operably linked regulatory elements; multiple genes, each with its owned set of operably linked regulatory elements; or multiple genes, regulated by the same set of regulatory elements. Intermediate transplacement constructs are often maintained in a replicon, such as an extrachromosomal element (*eg.* plasmids) capable of stable maintenance in a host, such as a bacterium. The replicon will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification.

Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 basepairs downstream from the ATT; see Luckow and Summers, *Virology* (1989) 17:31.

The plasmid usually also contains the polyhedrin polyadenylation signal (Miller et al. (1988) *Ann. Rev. Microbiol.*, 42:177) and a prokaryotic ampicillin-resistance (*amp*) gene and origin of replication for selection and propagation in *E. coli*.

Baculovirus transfer vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (5' to 3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A baculovirus transfer vector may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Expression may be either regulated or constitutive.

Structural genes, abundantly transcribed at late times in a viral infection cycle, provide particularly useful promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedron protein, Friesen et al., (1986) "The Regulation of Baculovirus Gene Expression," in: *The Molecular Biology of Baculoviruses* (ed. Walter Doerfler); EPO Publ. Nos. 127 839 and 155 476; and the gene encoding the p10 protein, Vlak et al., (1988), *J. Gen. Virol.* 69:765.

DNA encoding suitable signal sequences can be derived from genes for secreted insect or baculovirus proteins, such as the baculovirus polyhedrin gene (Carbonell et al. (1988) *Gene*, 73:409). Alternatively, since the signals for mammalian cell posttranslational modifications (such as signal peptide cleavage, proteolytic cleavage, and phosphorylation) appear to be recognized by insect cells, and the signals required for secretion and nuclear accumulation also appear to be conserved between the invertebrate cells and vertebrate cells, leaders of non-insect origin, such as those derived from genes encoding human  $\alpha$ -interferon, Maeda et al., (1985), *Nature* 315:592; human gastrin-releasing peptide, Lebacqz-Verheyden et al., (1988), *Molec. Cell. Biol.* 8:3129; human IL-2, Smith et al., (1985) *Proc. Nat'l Acad. Sci. USA*, 82:8404; mouse IL-3, (Miyajima et al., (1987) *Gene* 58:273; and human glucocerebrosidase, Martin et al. (1988) *DNA*, 7:99, can also be used to provide for secretion in insects.

A recombinant polypeptide or polyprotein may be expressed intracellularly or, if it is expressed with the proper regulatory sequences, it can be secreted. Good intracellular expression of nonfused foreign proteins usually requires heterologous genes that ideally have a short leader sequence containing suitable translation initiation signals preceding an ATG start signal. If desired, methionine at the N-terminus may be cleaved from the mature protein by *in vitro* incubation with cyanogen bromide.

Alternatively, recombinant polyproteins or proteins which are not naturally secreted can be secreted from the insect cell by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in insects. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the translocation of the protein into the endoplasmic reticulum.

After insertion of the DNA sequence and/or the gene encoding the expression product precursor of the protein, an insect cell host is co-transformed with the heterologous DNA of the transfer vector and the genomic DNA of wild type baculovirus -- usually by co-transfection. The promoter

and transcription termination sequence of the construct will usually comprise a 2-5kb section of the baculovirus genome. Methods for introducing heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summers and Smith *supra*; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* (1983) 3:2156; and Luckow and Summers (1989)). For example, the insertion can be into a gene such as the polyhedrin gene, by homologous double crossover recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. Miller et al., (1989), *Bioessays* 4:91. The DNA sequence, when cloned in place of the polyhedrin gene in the expression vector, is flanked both 5' and 3' by polyhedrin-specific sequences and is positioned downstream of the polyhedrin promoter.

- 10 The newly formed baculovirus expression vector is subsequently packaged into an infectious recombinant baculovirus. Homologous recombination occurs at low frequency (between about 1% and about 5%); thus, the majority of the virus produced after cotransfection is still wild-type virus. Therefore, a method is necessary to identify recombinant viruses. An advantage of the expression system is a visual screen allowing recombinant viruses to be distinguished. The polyhedrin protein, which is produced by the native virus, is produced at very high levels in the nuclei of infected cells at late times after viral infection. Accumulated polyhedrin protein forms occlusion bodies that also contain embedded particles. These occlusion bodies, up to 15  $\mu\text{m}$  in size, are highly refractile, giving them a bright shiny appearance that is readily visualized under the light microscope. Cells infected with recombinant viruses lack occlusion bodies. To distinguish recombinant virus from wild-type virus, the transfection supernatant is plaqued onto a monolayer of insect cells by techniques known to those skilled in the art. Namely, the plaques are screened under the light microscope for the presence (indicative of wild-type virus) or absence (indicative of recombinant virus) of occlusion bodies. "Current Protocols in Microbiology" Vol. 2 (Ausubel et al. eds) at 16.8 (Supp. 10, 1990); Summers and Smith, *supra*; Miller et al. (1989).
- 25 Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia*: *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni* (WO 89/046699; Carbonell et al., (1985) *J. Virol.* 56:153; Wright (1986) *Nature* 321:718; Smith et al., (1983) *Mol. Cell. Biol.* 3:2156; and see generally, Fraser, et al. (1989) *In Vitro Cell. Dev. Biol.* 25:225).
- 30

Cells and cell culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression system; cell culture technology is generally known to those skilled in the art. *See, eg. Summers and Smith supra.*

5 The modified insect cells may then be grown in an appropriate nutrient medium, which allows for stable maintenance of the plasmid(s) present in the modified insect host. Where the expression product gene is under inducible control, the host may be grown to high density, and expression induced. Alternatively, where expression is constitutive, the product will be continuously expressed into the medium and the nutrient medium must be continuously circulated, while removing the product of interest and augmenting depleted nutrients. The product may be purified by such techniques as  
10 chromatography, eg. HPLC, affinity chromatography, ion exchange chromatography, etc.; electrophoresis; density gradient centrifugation; solvent extraction, or the like. As appropriate, the product may be further purified, as required, so as to remove substantially any insect proteins which are also secreted in the medium or result from lysis of insect cells, so as to provide a product which is at least substantially free of host debris, eg. proteins, lipids and polysaccharides.

15 In order to obtain protein expression, recombinant host cells derived from the transformants are incubated under conditions which allow expression of the recombinant protein encoding sequence. These conditions will vary, dependent upon the host cell selected. However, the conditions are readily ascertainable to those of ordinary skill in the art, based upon what is known in the art.

### iii. Plant Systems

20 There are many plant cell culture and whole plant genetic expression systems known in the art. Exemplary plant cellular genetic expression systems include those described in patents, such as: US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, *Phytochemistry* 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the references described above in  
25 Vaulcombe et al., *Mol. Gen. Genet.* 209:33-40 (1987); Chandler et al., *Plant Molecular Biology* 3:407-418 (1984); Rogers, *J. Biol. Chem.* 260:3731-3738 (1985); Rothstein et al., *Gene* 55:353-356 (1987); Whittier et al., *Nucleic Acids Research* 15:2515-2535 (1987); Wirsal et al., *Molecular Microbiology* 3:3-14 (1989); Yu et al., *Gene* 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by  
30 gibberellic acid can be found in R.L. Jones and J. MacMillin, *Gibberellins*: in: *Advanced Plant Physiology*, Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited, London, pp. 21-52.

References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel and Hickey, *Proc. Natl. Acad. Sci.* 84:1337-1339 (1987)

Typically, using techniques known in the art, a desired polynucleotide sequence is inserted into an expression cassette comprising genetic regulatory elements designed for operation in plants. The expression cassette is inserted into a desired expression vector with companion sequences upstream and downstream from the expression cassette suitable for expression in a plant host. The companion sequences will be of plasmid or viral origin and provide necessary characteristics to the vector to permit the vectors to move DNA from an original cloning host, such as bacteria, to the desired plant host. The basic bacterial/plant vector construct will preferably provide a broad host range prokaryote replication origin; a prokaryote selectable marker; and, for *Agrobacterium* transformations, T DNA sequences for *Agrobacterium*-mediated transfer to plant chromosomes. Where the heterologous gene is not readily amenable to detection, the construct will preferably also have a selectable marker gene suitable for determining if a plant cell has been transformed. A general review of suitable markers, for example for the members of the grass family, is found in Wilkink and Dons, 1993, *Plant Mol. Biol. Repr.* 11(2):165-185.

Sequences suitable for permitting integration of the heterologous sequence into the plant genome are also recommended. These might include transposon sequences and the like for homologous recombination as well as Ti sequences which permit random insertion of a heterologous expression cassette into a plant genome. Suitable prokaryote selectable markers include resistance toward antibiotics such as ampicillin or tetracycline. Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art.

The nucleic acid molecules of the subject invention may be included into an expression cassette for expression of the protein(s) of interest. Usually, there will be only one expression cassette, although two or more are feasible. The recombinant expression cassette will contain in addition to the heterologous protein encoding sequence the following elements, a promoter region, plant 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes equipped with one, and a transcription and translation termination sequence. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

A heterologous coding sequence may be for any protein relating to the present invention. The sequence encoding the protein of interest will encode a signal peptide which allows processing and translocation of the protein, as appropriate, and will usually lack any sequence which might result in the binding of the desired protein of the invention to a membrane. Since, for the most part, the transcriptional initiation region will be for a gene which is expressed and translocated during germination, by employing the signal peptide which provides for translocation, one may also provide for translocation of the protein of interest. In this way, the protein(s) of interest will be translocated from the cells in which they are expressed and may be efficiently harvested. Typically secretion in seeds are across the aleurone or scutellar epithelium layer into the endosperm of the seed. While it is not required that the protein be secreted from the cells in which the protein is produced, this facilitates the isolation and purification of the recombinant protein.

Since the ultimate expression of the desired gene product will be in a eucaryotic cell it is desirable to determine whether any portion of the cloned gene contains sequences which will be processed out as introns by the host's splicosome machinery. If so, site-directed mutagenesis of the "intron" region may be conducted to prevent losing a portion of the genetic message as a false intron code, Reed and Maniatis, *Cell* 41:95-105, 1985.

The vector can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. Crossway, *Mol. Gen. Genet*, 202:179-185, 1985. The genetic material may also be transferred into the plant cell by using polyethylene glycol, Krens, et al., *Nature*, 296, 72-74, 1982. Another method of introduction of nucleic acid segments is high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface, Klein, et al., *Nature*, 327, 70-73, 1987 and Knudsen and Muller, 1991, *Planta*, 185:330-336 teaching particle bombardment of barley endosperm to create transgenic barley. Yet another method of introduction would be fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies, Fraley, et al., *Proc. Natl. Acad. Sci. USA*, 79, 1859-1863, 1982.

The vector may also be introduced into the plant cells by electroporation. (Fromm et al., *Proc. Natl. Acad. Sci. USA* 82:5824, 1985). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.



All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersion*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Hererocallis*, *Nemesia*, *Pelargonium*, *Panicum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browaalia*, *Glycine*, *Lolium*, *Zea*, *Triticum*, *Sorghum*, and *Datura*.

Means for regeneration vary from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted. Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

In some plant cell culture systems, the desired protein of the invention may be excreted or alternatively, the protein may be extracted from the whole plant. Where the desired protein of the invention is secreted into the medium, it may be collected. Alternatively, the embryos and embryoless-half seeds or other plant tissue may be mechanically disrupted to release any secreted protein between cells and tissues. The mixture may be suspended in a buffer solution to retrieve soluble proteins. Conventional protein isolation and purification methods will be then used to purify the recombinant protein. Parameters of time, temperature pH, oxygen, and volumes will be adjusted through routine methods to optimize expression and recovery of heterologous protein.

#### iv. Bacterial Systems

Bacterial expression techniques are known in the art. A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in *Escherichia coli* (*E. coli*) [Raibaud *et al.* (1984) *Annu. Rev. Genet.* 18:173]. Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (*lac*) [Chang *et al.* (1977) *Nature* 198:1056], and maltose. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan (*trp*) [Goeddel *et al.* (1980) *Nuc. Acids Res.* 8:4057; Yelverton *et al.* (1981) *Nucl. Acids Res.* 9:731; US patent 4,738,921; EP-A-0036776 and EP-A-0121775]. The g-lactamase (*bla*) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In *Interferon 3* (ed. I. Gresser)], bacteriophage lambda PL [Shimatake *et al.* (1981) *Nature* 292:128] and T5 [US patent 4,689,406] promoter systems also provide useful promoter sequences.

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [US patent 4,551,433]. For example, the *tac* promoter is a hybrid *trp-lac* promoter comprised of both *trp* promoter and *lac* operon sequences that is regulated by the *lac* repressor [Amann *et al.* (1983) *Gene* 25:167; de Boer *et al.* (1983) *Proc. Natl. Acad. Sci.* 80:21].

Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA polymerase/promoter system is an example of a coupled promoter system [Studier *et al.* (1986) *J. Mol. Biol.* 189:113; Tabor *et al.* (1985) *Proc Natl. Acad. Sci.* 82:1074]. In addition, a hybrid promoter can also be comprised of a bacteriophage promoter and an *E. coli* operator region (EPO-A-0 267 851).

In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of foreign genes in prokaryotes. In *E. coli*, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine *et al.* (1975) *Nature* 254:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' end of *E. coli* 16S rRNA [Steitz *et al.* (1979) "Genetic signals and nucleotide sequences in messenger RNA." In *Biological Regulation and Development: Gene Expression* (ed. R.F. Goldberger)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-binding site [Sambrook *et al.* (1989) "Expression of cloned genes in *Escherichia coli*." In *Molecular Cloning: A Laboratory Manual*].

A DNA molecule may be expressed intracellularly. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide or by either *in vivo* or *in vitro* incubation with a bacterial methionine N-terminal peptidase (EPO-A-0 219 237).

Fusion proteins provide an alternative to direct expression. Usually, a DNA sequence encoding the N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of a foreign gene and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the foreign gene [Nagai *et al.* (1984) *Nature* 309:810]. Fusion proteins can also be made with sequences from the *lacZ* [Jia *et al.* (1987) *Gene* 60:197], *trpE* [Allen *et al.* (1987) *J. Biotechnol.* 5:93; Makoff *et al.*

(1989) *J. Gen. Microbiol.* 135:11], and *Chey* [EP-A-0 324 647] genes. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (eg. ubiquitin specific processing-protease) to cleave the ubiquitin from the foreign protein. Through this method, native foreign protein can be isolated [Miller *et al.* (1989) *Bio/Technology* 7:698].

Alternatively, foreign proteins can also be secreted from the cell by creating chimeric DNA molecules that encode a fusion protein comprised of a signal peptide sequence fragment that provides for secretion of the foreign protein in bacteria [US patent 4,336,336]. The signal sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

DNA encoding suitable signal sequences can be derived from genes for secreted bacterial proteins, such as the *E. coli* outer membrane protein gene (*ompA*) [Masui *et al.* (1983), in: *Experimental Manipulation of Gene Expression*; Ghrayeb *et al.* (1984) *EMBO J.* 3:2437] and the *E. coli* alkaline phosphatase signal sequence (*phoA*) [Oka *et al.* (1985) *Proc. Natl. Acad. Sci.* 82:7212]. As an additional example, the signal sequence of the alpha-amylase gene from various *Bacillus* strains can be used to secrete heterologous proteins from *B. subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 244 042].

Usually, transcription termination sequences recognized by bacteria are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Transcription termination sequences frequently include DNA sequences of about 50 nucleotides capable of forming stem loop structures that aid in terminating transcription. Examples include transcription termination sequences derived from genes with strong promoters, such as the *trp* gene in *E. coli* as well as other biosynthetic genes.

Usually, the above described components, comprising a promoter, signal sequence (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal

element (eg. plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a prokaryotic host either for expression or for cloning and amplification. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably contain at least about 10, and more preferably at least about 20 plasmids. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host.

Alternatively, the expression constructs can be integrated into the bacterial genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to the bacterial chromosome that allows the vector to integrate. Integrations appear to result from recombinations between homologous DNA in the vector and the bacterial chromosome. For example, integrating vectors constructed with DNA from various *Bacillus* strains integrate into the *Bacillus* chromosome (EP-A- 0 127 328). Integrating vectors may also be comprised of bacteriophage or transposon sequences.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of bacterial strains that have been transformed. Selectable markers can be expressed in the bacterial host and may include genes which render bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin (neomycin), and tetracycline [Davies *et al.* (1978) *Annu. Rev. Microbiol.* 32:469]. Selectable markers may also include biosynthetic genes, such as those in the histidine, tryptophan, and leucine biosynthetic pathways.

Alternatively, some of the above described components can be put together in transformation vectors. Transformation vectors are usually comprised of a selectable market that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, *inter alia*, the following bacteria: *Bacillus subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541], *Escherichia coli* [Shimatake *et al.* (1981) *Nature* 292:128; Amann *et al.* (1985) *Gene* 40:183; Studier *et al.* (1986) *J. Mol. Biol.* 189:113; EP-A-0 036 776, EP-A-0 136 829 and EP-A-0 136 907],

*Streptococcus cremoris* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655]; *Streptococcus lividans* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655], *Streptomyces lividans* [US patent 4,745,056].

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with  $\text{CaCl}_2$  or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. See *eg.* [Masson *et al.* (1989) *FEMS Microbiol. Lett.* 60:273; Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, *Bacillus*], [Miller *et al.* (1988) *Proc. Natl. Acad. Sci.* 85:856; Wang *et al.* (1990) *J. Bacteriol.* 172:949, *Campylobacter*], [Cohen *et al.* (1973) *Proc. Natl. Acad. Sci.* 69:2110; Dower *et al.* (1988) *Nucleic Acids Res.* 16:6127; Kushner (1978) "An improved method for transformation of *Escherichia coli* with ColE1-derived plasmids. In *Genetic Engineering: Proceedings of the International Symposium on Genetic Engineering* (eds. H.W. Boyer and S. Nicosia); Mandel *et al.* (1970) *J. Mol. Biol.* 53:159; Taketo (1988) *Biochim. Biophys. Acta* 949:318; *Escherichia*], [Chassy *et al.* (1987) *FEMS Microbiol. Lett.* 44:173 *Lactobacillus*]; [Fiedler *et al.* (1988) *Anal. Biochem* 170:38, *Pseudomonas*]; [Augustin *et al.* (1990) *FEMS Microbiol. Lett.* 66:203, *Staphylococcus*], [Barany *et al.* (1980) *J. Bacteriol.* 144:698; Harlander (1987) "Transformation of *Streptococcus lactis* by electroporation, in: *Streptococcal Genetics* (ed. J. Ferretti and R. Curtiss III); Perry *et al.* (1981) *Infect. Immun.* 32:1295; Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655; Somkuti *et al.* (1987) *Proc. 4th Evr. Cong. Biotechnology* 1:412, *Streptococcus*].

#### v. Yeast Expression

Yeast expression systems are also known to one of ordinary skill in the art. A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the "TATA Box") and a transcription initiation site. A yeast promoter may also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible) expression. Constitutive expression occurs in the absence

of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.

Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples include alcohol dehydrogenase (ADH) (EP-A-0 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO-A-0 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences [Myanohara *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:1].

- 10 In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (US Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters
- 15 which consist of the regulatory sequences of either the *ADH2*, *GAL4*, *GAL10*, OR *PHO5* genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 164 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, *inter alia*, [Cohen *et al.* (1980) *Proc. Natl. Acad. Sci. USA*
- 20 77:1078; Henikoff *et al.* (1981) *Nature* 283:835; Hollenberg *et al.* (1981) *Curr. Topics Microbiol. Immunol.* 96:119; Hollenberg *et al.* (1979) "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast *Saccharomyces cerevisiae*," in: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon *et al.* (1980) *Gene* 11:163; Panthier *et al.* (1980) *Curr. Genet.* 2:109;].
- 25 A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See *eg.* EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*eg.* ubiquitin-specific processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (*eg.* WO88/024066).

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provide for secretion in yeast of the foreign protein. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell.

DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the yeast invertase gene (EP-A-0 012 873; JPO. 62,096,086) and the A-factor gene (US patent 4,588,684). Alternatively, leaders of non-yeast origin, such as an interferon leader, exist that also provide for secretion in yeast (EP-A-0 060 057).

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated alpha-factor leaders (usually about 25 to about 50 amino acid residues) (US Patents 4,546,083 and 4,870,008; EP-A-0 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alphafactor. (*eg.* see WO 89/02463.)



Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle vectors include YEp24 [Botstein *et al.* (1979) *Gene* 8:17-24], pCI/1 [Brake *et al.* (1984) *Proc. Natl. Acad. Sci USA* 81:4642-4646], and YRp17 [Stinchcomb *et al.* (1982) *J. Mol. Biol.* 158:157]. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Enter a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See eg. Brake *et al.*, *supra*.

Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome [Orr-Weaver *et al.* (1983) *Methods in Enzymol.* 101:228-245]. An integrating vector may be directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver *et al.*, *supra*. One or more expression construct may integrate, possibly affecting levels of recombinant protein produced [Rine *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:6750]. The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the

chromosome and flanking the expression construct in the vector, which can result in the stable integration of only the expression construct.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers may include biosynthetic genes that can be expressed in the yeast host, such as *ADE2*, *HIS4*, *LEU2*, *TRP1*, and *ALG7*, and the G418 resistance gene, which confer resistance in yeast cells to tunicamycin and G418, respectively. In addition, a suitable selectable marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of *CUP1* allows yeast to grow in the presence of copper ions [Butt *et al.* (1987) *Microbiol. Rev.* 51:351].

Alternatively, some of the above described components can be put together into transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, *inter alia*, the following yeasts: *Candida albicans* [Kurtz, *et al.* (1986) *Mol. Cell. Biol.* 6:142], *Candida maltosa* [Kunze, *et al.* (1985) *J. Basic Microbiol.* 25:141], *Hansenula polymorpha* [Gleeson, *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302], *Kluyveromyces fragilis* [Das, *et al.* (1984) *J. Bacteriol.* 158:1165], *Kluyveromyces lactis* [De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:737; Van den Berg *et al.* (1990) *Bio/Technology* 8:135], *Pichia guilliermondii* [Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141], *Pichia pastoris* [Cregg, *et al.* (1985) *Mol. Cell. Biol.* 5:3376; US Patent Nos. 4,837,148 and 4,929,555], *Saccharomyces cerevisiae* [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163], *Schizosaccharomyces pombe* [Beach and Nurse (1981) *Nature* 300:706], and *Yarrowia lipolytica* [Davidow, *et al.* (1985) *Curr. Genet.* 10:380471 Gaillardin, *et al.* (1985) *Curr. Genet.* 10:49].

Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. See *eg.* [Kurtz *et al.* (1986) *Mol. Cell. Biol.* 6:142; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; *Candida*];

[Gleeson *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302; Hansenula]; [Das *et al.* (1984) *J. Bacteriol.* 158:1165; De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:1165; Van den Berg *et al.* (1990) *Bio/Technology* 8:135; Kluyveromyces]; [Cregg *et al.* (1985) *Mol. Cell. Biol.* 5:3376; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; US Patent  
5 Nos. 4,837,148 and 4,929,555; Pichia]; [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163 *Saccharomyces*]; [Beach and Nurse (1981) *Nature* 300:706; *Schizosaccharomyces*]; [Davidow *et al.* (1985) *Curr. Genet.* 10:39; Gaillardin *et al.* (1985) *Curr. Genet.* 10:49; Yarrowia].

### Antibodies

10 As used herein, the term "antibody" refers to a polypeptide or group of polypeptides composed of at least one antibody combining site. An "antibody combining site" is the three-dimensional binding space with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows a binding of the antibody with the antigen. "Antibody" includes, for example, vertebrate antibodies, hybrid antibodies, chimeric antibodies, humanised  
15 antibodies, altered antibodies, univalent antibodies, Fab proteins, and single domain antibodies.

Antibodies against the proteins of the invention are useful for affinity chromatography, immunoassays, and distinguishing/identifying Neisserial proteins.

Antibodies to the proteins of the invention, both polyclonal and monoclonal, may be prepared by conventional methods. In general, the protein is first used to immunize a suitable animal, preferably  
20 a mouse, rat, rabbit or goat. Rabbits and goats are preferred for the preparation of polyclonal sera due to the volume of serum obtainable, and the availability of labeled anti-rabbit and anti-goat antibodies. Immunization is generally performed by mixing or emulsifying the protein in saline, preferably in an adjuvant such as Freund's complete adjuvant, and injecting the mixture or emulsion parenterally (generally subcutaneously or intramuscularly). A dose of 50-200 µg/injection  
25 is typically sufficient. Immunization is generally boosted 2-6 weeks later with one or more injections of the protein in saline, preferably using Freund's incomplete adjuvant. One may alternatively generate antibodies by *in vitro* immunization using methods known in the art, which for the purposes of this invention is considered equivalent to *in vivo* immunization. Polyclonal antisera is obtained by bleeding the immunized animal into a glass or plastic container, incubating  
30 the blood at 25°C for one hour, followed by incubating at 4°C for 2-18 hours. The serum is

recovered by centrifugation (eg. 1,000g for 10 minutes). About 20-50 ml per bleed may be obtained from rabbits.

Monoclonal antibodies are prepared using the standard method of Kohler & Milstein [*Nature* (1975) 256:495-96], or a modification thereof. Typically, a mouse or rat is immunized as described  
5 above. However, rather than bleeding the animal to extract serum, the spleen (and optionally several large lymph nodes) is removed and dissociated into single cells. If desired, the spleen cells may be screened (after removal of nonspecifically adherent cells) by applying a cell suspension to a plate or well coated with the protein antigen. B-cells expressing membrane-bound immunoglobulin specific for the antigen bind to the plate, and are not rinsed away with the rest of  
10 the suspension. Resulting B-cells, or all dissociated spleen cells, are then induced to fuse with myeloma cells to form hybridomas, and are cultured in a selective medium (eg. hypoxanthine, aminopterin, thymidine medium, "HAT"). The resulting hybridomas are plated by limiting dilution, and are assayed for the production of antibodies which bind specifically to the immunizing antigen (and which do not bind to unrelated antigens). The selected MAb-secreting hybridomas are then  
15 cultured either *in vitro* (eg. in tissue culture bottles or hollow fiber reactors), or *in vivo* (as ascites in mice).

If desired, the antibodies (whether polyclonal or monoclonal) may be labeled using conventional techniques. Suitable labels include fluorophores, chromophores, radioactive atoms (particularly  $^{32}\text{P}$  and  $^{125}\text{I}$ ), electron-dense reagents, enzymes, and ligands having specific binding partners. Enzymes  
20 are typically detected by their activity. For example, horseradish peroxidase is usually detected by its ability to convert 3,3',5,5'-tetramethylbenzidine (TMB) to a blue pigment, quantifiable with a spectrophotometer. "Specific binding partner" refers to a protein capable of binding a ligand molecule with high specificity, as for example in the case of an antigen and a monoclonal antibody specific therefor. Other specific binding partners include biotin and avidin or streptavidin, IgG and protein A,  
25 and the numerous receptor-ligand couples known in the art. It should be understood that the above description is not meant to categorize the various labels into distinct classes, as the same label may serve in several different modes. For example,  $^{125}\text{I}$  may serve as a radioactive label or as an electron-dense reagent. HRP may serve as enzyme or as antigen for a MAb. Further, one may combine various labels for desired effect. For example, MAbs and avidin also require labels in the practice of  
30 this invention: thus, one might label a MAb with biotin, and detect its presence with avidin labeled with  $^{125}\text{I}$ , or with an anti-biotin MAb labeled with HRP. Other permutations and possibilities will be

readily apparent to those of ordinary skill in the art, and are considered as equivalents within the scope of the instant invention.

### Pharmaceutical Compositions

Pharmaceutical compositions can comprise either polypeptides, antibodies, or nucleic acid of the invention. The pharmaceutical compositions will comprise a therapeutically effective amount of  
5 either polypeptides, antibodies, or polynucleotides of the claimed invention.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or  
10 antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation can be determined by routine  
15 experimentation and is within the judgement of the clinician.

For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such  
20 as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus  
25 particles. Such carriers are well known to those of ordinary skill in the art.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in Remington's Pharmaceutical Sciences (Mack  
30 Pub. Co., N.J. 1991).

- Pharmaceutically acceptable carriers in therapeutic compositions may contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier.

### Delivery Methods

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

- 10 Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (eg. see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

### Vaccines

Vaccines according to the invention may either be prophylactic (*ie.* to prevent infection) or therapeutic (*ie.* to treat disease after infection).

- Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with "pharmaceutically acceptable carriers," which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, etc. pathogens.

- Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc;
- 30 (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents

such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59™ (WO 90/14837; Chapter 10 in *Vaccine design: the subunit and adjuvant approach*, eds. Powell & Newman, Plenum Press 1995), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (eg. IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, etc.), interferons (eg. gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc; and (6) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59™ are preferred.

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-*sn*-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), etc.

The immunogenic compositions (eg. the immunising antigen/immunogen/polypeptide/protein/nucleic acid, pharmaceutically acceptable carrier, and adjuvant) typically will contain diluents, such as water, saline, glycerol, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.

Typically, the immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above under pharmaceutically acceptable carriers.

Immunogenic compositions used as vaccines comprise an immunologically effective amount of the antigenic or immunogenic polypeptides, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated (*eg.* nonhuman primate, primate, *etc.*), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

The immunogenic compositions are conventionally administered parenterally, *eg.* by injection, either subcutaneously, intramuscularly, or transdermally/transcutaneously (*eg.* WO98/20734). Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be administered in conjunction with other immunoregulatory agents.

As an alternative to protein-based vaccines, DNA vaccination may be employed [*eg.* Robinson & Torres (1997) *Seminars in Immunology* 9:271-283; Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648; see later herein].

## 20 Gene Delivery Vehicles

Gene therapy vehicles for delivery of constructs including a coding sequence of a therapeutic of the invention, to be delivered to the mammal for expression in the mammal, can be administered either locally or systemically. These constructs can utilize viral or non-viral vector approaches in *in vivo* or *ex vivo* modality. Expression of such coding sequence can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence *in vivo* can be either constitutive or regulated.

The invention includes gene delivery vehicles capable of expressing the contemplated nucleic acid sequences. The gene delivery vehicle is preferably a viral vector and, more preferably, a retroviral, adenoviral, adeno-associated viral (AAV), herpes viral, or alphavirus vector. The viral vector can also be an astrovirus, coronavirus, orthomyxovirus, papovavirus, paramyxovirus, parvovirus,



picornavirus, poxvirus, or togavirus viral vector. See generally, Jolly (1994) *Cancer Gene Therapy* 1:51-64; Kimura (1994) *Human Gene Therapy* 5:845-852; Connelly (1995) *Human Gene Therapy* 6:185-193; and Kaplitt (1994) *Nature Genetics* 6:148-153.

5 Retroviral vectors are well known in the art and we contemplate that any retroviral gene therapy vector is employable in the invention, including B, C and D type retroviruses, xenotropic retroviruses (for example, NZB-X1, NZB-X2 and NZB9-1 (see O'Neill (1985) *J. Virol.* 53:160) polytropic retroviruses eg. MCF and MCF-MLV (see Kelly (1983) *J. Virol.* 45:291), spumaviruses and lentiviruses. See RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985.

10 Portions of the retroviral gene therapy vector may be derived from different retroviruses. For example, retrovector LTRs may be derived from a Murine Sarcoma Virus, a tRNA binding site from a Rous Sarcoma Virus, a packaging signal from a Murine Leukemia Virus, and an origin of second strand synthesis from an Avian Leukosis Virus.

15 These recombinant retroviral vectors may be used to generate transduction competent retroviral vector particles by introducing them into appropriate packaging cell lines (see US patent 5,591,624). Retrovirus vectors can be constructed for site-specific integration into host cell DNA by incorporation of a chimeric integrase enzyme into the retroviral particle (see WO96/37626). It is preferable that the recombinant viral vector is a replication defective recombinant virus.

20 Packaging cell lines suitable for use with the above-described retrovirus vectors are well known in the art, are readily prepared (see WO95/30763 and WO92/05266), and can be used to create producer cell lines (also termed vector cell lines or "VCLs") for the production of recombinant vector particles. Preferably, the packaging cell lines are made from human parent cells (eg. HT1080 cells) or mink parent cell lines, which eliminates inactivation in human serum.

25 Preferred retroviruses for the construction of retroviral gene therapy vectors include Avian Leukosis Virus, Bovine Leukemia Virus, Murine Leukemia Virus, Mink-Cell Focus-Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis Virus and Rous Sarcoma Virus. Particularly preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley and Rowe (1976) *J Virol* 19:19-25), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Graffi, Gross (ATCC No. VR-590), Kirsten, Harvey Sarcoma Virus and Rauscher (ATCC No. VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190). Such retroviruses may be obtained from depositories or

collections such as the American Type Culture Collection ("ATCC") in Rockville, Maryland or isolated from known sources using commonly available techniques.

Exemplary known retroviral gene therapy vectors employable in this invention include those described in patent applications GB2200651, EP0415731, EP0345242, EP0334301, WO89/02468; 5 WO89/05349, WO89/09271, WO90/02806, WO90/07936, WO94/03622, WO93/25698, WO93/25234, WO93/11230, WO93/10218, WO91/02805, WO91/02825, WO95/07994, US 5,219,740, US 4,405,712, US 4,861,719, US 4,980,289, US 4,777,127, US 5,591,624. See also Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53 (1993) 83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 10 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

Human adenoviral gene therapy vectors are also known in the art and employable in this invention. See, for example, Berkner (1988) *Biotechniques* 6:616 and Rosenfeld (1991) *Science* 252:431, and WO93/07283, WO93/06223, and WO93/07282. Exemplary known adenoviral gene therapy vectors 15 employable in this invention include those described in the above referenced documents and in WO94/12649, WO93/03769, WO93/19191, WO94/28938, WO95/11984, WO95/00655, WO95/27071, WO95/29993, WO95/34671, WO96/05320, WO94/08026, WO94/11506, WO93/06223, WO94/24299, WO95/14102, WO95/24297, WO95/02697, WO94/28152, WO94/24299, WO95/09241, WO95/25807, WO95/05835, WO94/18922 and WO95/09654.

20 Alternatively, administration of DNA linked to killed adenovirus as described in Curiel (1992) *Hum. Gene Ther.* 3:147-154 may be employed. The gene delivery vehicles of the invention also include adenovirus associated virus (AAV) vectors. Leading and preferred examples of such vectors for use in this invention are the AAV-2 based vectors disclosed in Srivastava, WO93/09239. Most preferred AAV vectors comprise the two AAV inverted terminal repeats in 25 which the native D-sequences are modified by substitution of nucleotides, such that at least 5 native nucleotides and up to 18 native nucleotides, preferably at least 10 native nucleotides up to 18 native nucleotides, most preferably 10 native nucleotides are retained and the remaining nucleotides of the D-sequence are deleted or replaced with non-native nucleotides. The native D-sequences of the AAV inverted terminal repeats are sequences of 20 consecutive nucleotides in each AAV inverted 30 terminal repeat (*ie.* there is one sequence at each end) which are not involved in HP formation. The non-native replacement nucleotide may be any nucleotide other than the nucleotide found in the

native D-sequence in the same position. Other employable exemplary AAV vectors are pWP-19, pWN-1, both of which are disclosed in Nahreini (1993) *Gene* 124:257-262. Another example of such an AAV vector is psub201 (see Samulski (1987) *J. Virol.* 61:3096). Another exemplary AAV vector is the Double-D ITR vector. Construction of the Double-D ITR vector is disclosed in US Patent 5,478,745. Still other vectors are those disclosed in Carter US Patent 4,797,368 and Muzyczka US Patent 5,139,941, Chartejee US Patent 5,474,935, and Kotin WO94/288157. Yet a further example of an AAV vector employable in this invention is SSV9AFABTKneo, which contains the AFP enhancer and albumin promoter and directs expression predominantly in the liver. Its structure and construction are disclosed in Su (1996) *Human Gene Therapy* 7:463-470.

5 Additional AAV gene therapy vectors are described in US 5,354,678, US 5,173,414, US 5,139,941, and US 5,252,479.

The gene therapy vectors of the invention also include herpes vectors. Leading and preferred examples are herpes simplex virus vectors containing a sequence encoding a thymidine kinase polypeptide such as those disclosed in US 5,288,641 and EP0176170 (Roizman). Additional exemplary herpes simplex virus vectors include HFEM/ICP6-LacZ disclosed in WO95/04139 (Wistar Institute), pHSVlac described in Geller (1988) *Science* 241:1667-1669 and in WO90/09441 and WO92/07945, HSV Us3::pgC-lacZ described in Fink (1992) *Human Gene Therapy* 3:11-19 and HSV 7134, 2 RH 105 and GAL4 described in EP 0453242 (Breakefield), and those deposited with the ATCC as accession numbers ATCC VR-977 and ATCC VR-260.

Also contemplated are alpha virus gene therapy vectors that can be employed in this invention. Preferred alpha virus vectors are Sindbis viruses vectors. Togaviruses, Semliki Forest virus (ATCC VR-67; ATCC VR-1247), Middleberg virus (ATCC VR-370), Ross River virus (ATCC VR-373; ATCC VR-1246), Venezuelan equine encephalitis virus (ATCC VR923; ATCC VR-1250; ATCC VR-1249; ATCC VR-532), and those described in US patents 5,091,309, 5,217,879, and WO92/10578. More particularly, those alpha virus vectors described in US Serial No. 08/405,627, filed March 15, 1995, WO94/21792, WO92/10578, WO95/07994, US 5,091,309 and US 5,217,879 are employable. Such alpha viruses may be obtained from depositories or collections such as the ATCC in Rockville, Maryland or isolated from known sources using commonly available techniques. Preferably, alphavirus vectors with reduced cytotoxicity are used (see USSN 08/679640).

DNA vector systems such as eukaryotic layered expression systems are also useful for expressing the nucleic acids of the invention. See WO95/07994 for a detailed description of eukaryotic layered expression systems. Preferably, the eukaryotic layered expression systems of the invention are derived from alphavirus vectors and most preferably from Sindbis viral vectors.

- 5 Other viral vectors suitable for use in the present invention include those derived from poliovirus, for example ATCC VR-58 and those described in Evans, *Nature* 339 (1989) 385 and Sabin (1973) *J. Biol. Standardization* 1:115; rhinovirus, for example ATCC VR-1110 and those described in Arnold (1990) *J Cell Biochem* L401; pox viruses such as canary pox virus or vaccinia virus, for example ATCC VR-111 and ATCC VR-2010 and those described in Fisher-Hoch (1989) *Proc Natl Acad Sci* 86:317; Flexner (1989) *Ann NY Acad Sci* 569:86, Flexner (1990) *Vaccine* 8:17; in US 4,603,112 and US 4,769,330 and WO89/01973; SV40 virus, for example ATCC VR-305 and those described in Mulligan (1979) *Nature* 277:108 and Madzak (1992) *J Gen Virol* 73:1533; influenza virus, for example ATCC VR-797 and recombinant influenza viruses made employing reverse genetics techniques as described in US 5,166,057 and in Enami (1990) *Proc Natl Acad Sci* 87:3802-3805; Enami & Palese (1991) *J Virol* 65:2711-2713 and Luytjes (1989) *Cell* 59:110, (see also McMichael (1983) *NEJ Med* 309:13, and Yap (1978) *Nature* 273:238 and *Nature* (1979) 277:108); human immunodeficiency virus as described in EP-0386882 and in Buchschacher (1992) *J. Virol.* 66:2731; measles virus, for example ATCC VR-67 and VR-1247 and those described in EP-0440219; Aura virus, for example ATCC VR-368; Bebaru virus, for example ATCC VR-600 and ATCC VR-1240; Cabassou virus, for example ATCC VR-922; Chikungunya virus, for example ATCC VR-64 and ATCC VR-1241; Fort Morgan Virus, for example ATCC VR-924; Getah virus, for example ATCC VR-369 and ATCC VR-1243; Kyzylagach virus, for example ATCC VR-927; Mayaro virus, for example ATCC VR-66; Mucambo virus, for example ATCC VR-580 and ATCC VR-1244; Ndumu virus, for example ATCC VR-371; Pixuna virus, for example ATCC VR-372 and ATCC VR-1245; Tonate virus, for example ATCC VR-925; Trinit virus, for example ATCC VR-469; Una virus, for example ATCC VR-374; Whataroa virus, for example ATCC VR-926; Y-62-33 virus, for example ATCC VR-375; O'Nyong virus, Eastern encephalitis virus, for example ATCC VR-65 and ATCC VR-1242; Western encephalitis virus, for example ATCC VR-70, ATCC VR-1251, ATCC VR-622 and ATCC VR-1252; and coronavirus, for example ATCC VR-740 and those described in Hamre (1966) *Proc Soc Exp Biol Med* 121:190.

Delivery of the compositions of this invention into cells is not limited to the above mentioned viral vectors. Other delivery methods and media may be employed such as, for example, nucleic acid

expression vectors, polycationic condensed DNA linked or unlinked to killed adenovirus alone, for example see US Serial No. 08/366,787, filed December 30, 1994 and Curiel (1992) *Hum Gene Ther* 3:147-154 ligand linked DNA, for example see Wu (1989) *J Biol Chem* 264:16985-16987, eucaryotic cell delivery vehicles cells, for example see US Serial No.08/240,030, filed May 9, 1994, and US Serial No. 08/404,796, deposition of photopolymerized hydrogel materials, hand-held gene transfer particle gun, as described in US Patent 5,149,655, ionizing radiation as described in US5,206,152 and in WO92/11033, nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip (1994) *Mol Cell Biol* 14:2411-2418 and in Woffendin (1994) *Proc Natl Acad Sci* 91:1581-1585.

- 10 Particle mediated gene transfer may be employed, for example see US Serial No. 60/023,867. Briefly, the sequence can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, as described in Wu & Wu (1987) *J. Biol. Chem.* 262:4429-4432, insulin as described in Hucked (1990) *Biochem Pharmacol* 40:253-263, galactose as described in Plank (1992) *Bioconjugate Chem* 3:533-539, lactose or transferrin.

Naked DNA may also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and US 5,580,859. Uptake efficiency may be improved using biodegradable latex beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the beads. The method may be improved further by treatment of the beads to increase hydrophobicity and thereby facilitate disruption of the endosome and release of the DNA into the cytoplasm.

Liposomes that can act as gene delivery vehicles are described in US 5,422,120, WO95/13796, WO94/23697, WO91/14445 and EP-524,968. As described in USSN. 60/023,867, on non-viral delivery, the nucleic acid sequences encoding a polypeptide can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then be incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, insulin, galactose, lactose, or transferrin. Other delivery systems include the use of liposomes to encapsulate DNA comprising the gene under the control of a variety of tissue-specific or ubiquitously-active promoters. Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al* (1994) *Proc. Natl. Acad. Sci. USA*

91(24):11581-11585. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun, as described in US 5,149,655; use of ionizing radiation for activating transferred gene, as described in US 5,206,152 and WO92/11033

Exemplary liposome and polycationic gene delivery vehicles are those described in US 5,422,120 and 4,762,915; in WO 95/13796; WO94/23697; and WO91/14445; in EP-0524968; and in Stryer, *Biochemistry*, pages 236-240 (1975) W.H. Freeman, San Francisco; Szoka (1980) *Biochem Biophys Acta* 600:1; Bayer (1979) *Biochem Biophys Acta* 550:464; Rivnay (1987) *Meth Enzymol* 149:119; Wang (1987) *Proc Natl Acad Sci* 84:7851; Plant (1989) *Anal Biochem* 176:420.

A polynucleotide composition can comprises therapeutically effective amount of a gene therapy vehicle, as the term is defined above. For purposes of the present invention, an effective dose will be from about 0.01 mg/kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

#### Delivery Methods

Once formulated, the polynucleotide compositions of the invention can be administered (1) directly to the subject; (2) delivered *ex vivo*, to cells derived from the subject; or (3) *in vitro* for expression of recombinant proteins. The subjects to be treated can be mammals or birds. Also, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (*eg.* see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in *eg.* WO93/14778. Examples of cells useful in *ex vivo* applications include, for example, stem cells, particularly hematopoietic, lymph cells, macrophages, dendritic cells, or tumor cells.

Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be accomplished by the following procedures, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

#### Polynucleotide and polypeptide pharmaceutical compositions

In addition to the pharmaceutically acceptable carriers and salts described above, the following additional agents can be used with polynucleotide and/or polypeptide compositions.

##### A. Polypeptides

One example are polypeptides which include, without limitation: asioloorosomuroid (ASOR); transferrin; asialoglycoproteins; antibodies; antibody fragments; ferritin; interleukins; interferons, granulocyte, macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor and erythropoietin. Viral antigens, such as envelope proteins, can also be used. Also, proteins from other invasive organisms, such as the 17 amino acid peptide from the circumsporozoite protein of plasmodium falciparum known as RII.

##### B. Hormones, Vitamins, etc.

Other groups that can be included are, for example: hormones, steroids, androgens, estrogens, thyroid hormone, or vitamins, folic acid.

##### C. Polyalkylenes, Polysaccharides, etc.

Also, polyalkylene glycol can be included with the desired polynucleotides/polypeptides. In a preferred embodiment, the polyalkylene glycol is polyethylene glycol. In addition, mono-, di-, or polysaccharides can be included. In a preferred embodiment of this aspect, the polysaccharide is dextran or DEAE-dextran. Also, chitosan and poly(lactide-co-glycolide)

##### D. Lipids, and Liposomes

The desired polynucleotide/polypeptide can also be encapsulated in lipids or packaged in liposomes prior to delivery to the subject or to cells derived therefrom.

Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed polynucleotide to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the

use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight (1991) *Biochim. Biophys. Acta*. 1097:1-17; Straubinger (1983) *Meth. Enzymol.* 101:512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations. Cationic liposomes have been shown to mediate intracellular delivery of plasmid DNA (Felgner (1987) *Proc. Natl. Acad. Sci. USA* 84:7413-7416); mRNA (Malone (1989) *Proc. Natl. Acad. Sci. USA* 86:6077-6081); and purified transcription factors (Debs (1990) *J. Biol. Chem.* 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N[1-2,3-dioleoyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See, also, Felgner *supra*). Other commercially available liposomes include transfectace (DDAB/DOPE) and DOTAP/DOPE (Boehringer). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, *eg.* Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; WO90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphosphatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilammellar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See *eg.* Straubinger (1983) *Meth. Immunol.* 101:512-527; Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; Papahadjopoulos (1975) *Biochim. Biophys. Acta* 394:483; Wilson (1979) *Cell* 17:77; Deamer & Bangham (1976) *Biochim. Biophys. Acta* 443:629; Ostro (1977) *Biochem. Biophys. Res. Commun.* 76:836; Fraley (1979) *Proc. Natl. Acad. Sci. USA* 76:3348; Enoch & Strittmatter (1979) *Proc. Natl. Acad. Sci. USA* 76:145; Fraley (1980) *J. Biol. Chem.* (1980) 255:10431; Szoka & Papahadjopoulos (1978) *Proc. Natl. Acad. Sci. USA* 75:145; and Schaefer-Ridder (1982) *Science* 215:166.



### E.Lipoproteins

In addition, lipoproteins can be included with the polynucleotide/polypeptide to be delivered. Examples of lipoproteins to be utilized include: chylomicrons, HDL, IDL, LDL, and VLDL. Mutants, fragments, or fusions of these proteins can also be used. Also, modifications of naturally occurring lipoproteins can be used, such as acetylated LDL. These lipoproteins can target the delivery of polynucleotides to cells expressing lipoprotein receptors. Preferably, if lipoproteins are including with the polynucleotide to be delivered, no other targeting ligand is included in the composition.

Naturally occurring lipoproteins comprise a lipid and a protein portion. The protein portion are known as apoproteins. At the present, apoproteins A, B, C, D, and E have been isolated and identified. At least two of these contain several proteins, designated by Roman numerals, AI, AII, AIV; CI, CII, CIII.

A lipoprotein can comprise more than one apoprotein. For example, naturally occurring chylomicrons comprises of A, B, C, and E, over time these lipoproteins lose A and acquire C and E apoproteins. VLDL comprises A, B, C, and E apoproteins, LDL comprises apoprotein B; and HDL comprises apoproteins A, C, and E.

The amino acid of these apoproteins are known and are described in, for example, Breslow (1985) *Annu Rev. Biochem* 54:699; Law (1986) *Adv. Exp Med. Biol.* 151:162; Chen (1986) *J Biol Chem* 261:12918; Kane (1980) *Proc Natl Acad Sci USA* 77:2465; and Utermann (1984) *Hum Genet* 65:232.

Lipoproteins contain a variety of lipids including, triglycerides, cholesterol (free and esters), and phospholipids. The composition of the lipids varies in naturally occurring lipoproteins. For example, chylomicrons comprise mainly triglycerides. A more detailed description of the lipid content of naturally occurring lipoproteins can be found, for example, in *Meth. Enzymol.* 128 (1986). The composition of the lipids are chosen to aid in conformation of the apoprotein for receptor binding activity. The composition of lipids can also be chosen to facilitate hydrophobic interaction and association with the polynucleotide binding molecule.

Naturally occurring lipoproteins can be isolated from serum by ultracentrifugation, for instance. Such methods are described in *Meth. Enzymol. (supra)*; Pitas (1980) *J. Biochem.* 255:5454-5460 and Mahey (1979) *J Clin. Invest* 64:743-750. Lipoproteins can also be produced by *in vitro* or recombinant methods by expression of the apoprotein genes in a desired host cell. See, for example, Atkinson (1986) *Annu Rev Biophys Chem* 15:403 and Radding (1958) *Biochim Biophys Acta* 30:

443. Lipoproteins can also be purchased from commercial suppliers, such as Biomedical Technologies, Inc., Stoughton, Massachusetts, USA. Further description of lipoproteins can be found in Zuckermann *et al.* PCT/US97/14465.

#### F. Polycationic Agents

- 5 Polycationic agents can be included, with or without lipoprotein, in a composition with the desired polynucleotide/polypeptide to be delivered.

Polycationic agents, typically, exhibit a net positive charge at physiological relevant pH and are capable of neutralizing the electrical charge of nucleic acids to facilitate delivery to a desired location. These agents have both in vitro, ex vivo, and in vivo applications. Polycationic agents can  
10 be used to deliver nucleic acids to a living subject either intramuscularly, subcutaneously, etc.

The following are examples of useful polypeptides as polycationic agents: polylysine, polyarginine, polyornithine, and protamine. Other examples include histones, protamines, human serum albumin, DNA binding proteins, non-histone chromosomal proteins, coat proteins from DNA viruses, such as (X174, transcriptional factors also contain domains that bind DNA and therefore may be useful  
15 as nucleic acid condensing agents. Briefly, transcriptional factors such as C/CEBP, c-jun, c-fos, AP-1, AP-2, AP-3, CPF, Prot-1, Sp-1, Oct-1, Oct-2, CREP, and TFIID contain basic domains that bind DNA sequences.

Organic polycationic agents include: spermine, spermidine, and putrescine.

The dimensions and of the physical properties of a polycationic agent can be extrapolated from the  
20 list above, to construct other polypeptide polycationic agents or to produce synthetic polycationic agents.

Synthetic polycationic agents which are useful include, for example, DEAE-dextran, polybrene. Lipofectin™, and lipofectAMINE™ are monomers that form polycationic complexes when combined with polynucleotides/polypeptides.

#### 25 Immunodiagnostic Assays

Neisserial antigens of the invention can be used in immunoassays to detect antibody levels (or, conversely, anti-Neisserial antibodies can be used to detect antigen levels). Immunoassays based on well defined, recombinant antigens can be developed to replace invasive diagnostics methods. Antibodies to Neisserial proteins within biological samples, including for example, blood or serum

samples, can be detected. Design of the immunoassays is subject to a great deal of variation, and a variety of these are known in the art. Protocols for the immunoassay may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed by packaging the appropriate materials, including the compositions of the invention, in suitable containers, along with the remaining reagents and materials (for example, suitable buffers, salt solutions, etc.) required for the conduct of the assay, as well as suitable set of assay instructions.

#### Nucleic Acid Hybridisation

"Hybridization" refers to the association of two nucleic acid sequences to one another by hydrogen bonding. Typically, one sequence will be fixed to a solid support and the other will be free in solution. Then, the two sequences will be placed in contact with one another under conditions that favor hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase sequence to the solid support (Denhardt's reagent or BLOTTO); concentration of the sequences; use of compounds to increase the rate of association of sequences (dextran sulfate or polyethylene glycol); and the stringency of the washing conditions following hybridization. See Sambrook *et al.* [*supra*] Volume 2, chapter 9, pages 9.47 to 9.57.

"Stringency" refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 120 to 200°C below the calculated  $T_m$  of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook *et al.* at page 9.50.

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The

total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1 µg for a plasmid or phage digest to  $10^{-9}$  to  $10^{-8}$  g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of probes can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 µg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of  $10^8$  cpm/µg. For a single-copy mammalian gene a conservative approach would start with 10 µg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than  $10^8$  cpm/µg, resulting in an exposure time of ~24 hours.

- 10 Several factors can affect the melting temperature ( $T_m$ ) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

$$T_m = 81 + 16.6(\log_{10} C_i) + 0.4[\%(G + C)] - 0.6(\%\text{formamide}) - 600/n - 1.5(\%\text{mismatch}).$$

where  $C_i$  is the salt concentration (monovalent ions) and  $n$  is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) *Anal. Biochem.* 138: 267-284).

- 20 In designing a hybridization experiment, some factors affecting nucleic acid hybridization can be conveniently altered. The temperature of the hybridization and washes and the salt concentration during the washes are the simplest to adjust. As the temperature of the hybridization increases (*ie.* stringency), it becomes less likely for hybridization to occur between strands that are nonhomologous, and as a result, background decreases. If the radiolabeled probe is not completely homologous with the immobilized fragment (as is frequently the case in gene family and interspecies hybridization experiments), the hybridization temperature must be reduced, and background will increase. The temperature of the washes affects the intensity of the hybridizing band and the degree of background in a similar manner. The stringency of the washes is also increased with decreasing salt concentrations.

- 30 In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with is 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology,

and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to start with both hybridization and wash conditions which are nonstringent. If non-specific bands or high background are observed after autoradiography, the filter can be washed at high stringency and reexposed. If the time required for exposure makes this approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

#### Nucleic Acid Probe Assays

Methods such as PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes according to the invention can determine the presence of cDNA or mRNA. A probe is said to "hybridize" with a sequence of the invention if it can form a duplex or double stranded complex, which is stable enough to be detected.

The nucleic acid probes will hybridize to the Neisserial nucleotide sequences of the invention (including both sense and antisense strands). Though many different nucleotide sequences will encode the amino acid sequence, the native Neisserial sequence is preferred because it is the actual sequence present in cells. mRNA represents a coding sequence and so a probe should be complementary to the coding sequence; single-stranded cDNA is complementary to mRNA, and so a cDNA probe should be complementary to the non-coding sequence.

The probe sequence need not be identical to the Neisserial sequence (or its complement) — some variation in the sequence and length can lead to increased assay sensitivity if the nucleic acid probe can form a duplex with target nucleotides, which can be detected. Also, the nucleic acid probe can include additional nucleotides to stabilize the formed duplex. Additional Neisserial sequence may also be helpful as a label to detect the formed duplex. For example, a non-complementary nucleotide sequence may be attached to the 5' end of the probe, with the remainder of the probe sequence being complementary to a Neisserial sequence. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the a Neisserial sequence in order to hybridize therewith and thereby form a duplex which can be detected.

The exact length and sequence of the probe will depend on the hybridization conditions, such as temperature, salt condition and the like. For example, for diagnostic applications, depending on the

complexity of the analyte sequence, the nucleic acid probe typically contains at least 10-20 nucleotides, preferably 15-25, and more preferably at least 30 nucleotides, although it may be shorter than this. Short primers generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

- 5 Probes may be produced by synthetic procedures, such as the triester method of Matteucci *et al.* [*J. Am. Chem. Soc.* (1981) 103:3185], or according to Urdea *et al.* [*Proc. Natl. Acad. Sci. USA* (1983) 80: 7461], or using commercially available automated oligonucleotide synthesizers.

The chemical nature of the probe can be selected according to preference. For certain applications, DNA or RNA are appropriate. For other applications, modifications may be incorporated *eg.*

- 10 backbone modifications, such as phosphorothioates or methylphosphonates, can be used to increase *in vivo* half-life, alter RNA affinity, increase nuclease resistance *etc.* [*eg.* see Agrawal & Iyer (1995) *Curr Opin Biotechnol* 6:12-19; Agrawal (1996) *TIBTECH* 14:376-387]; analogues such as peptide nucleic acids may also be used [*eg.* see Corey (1997) *TIBTECH* 15:224-229; Buchardt *et al.* (1993) *TIBTECH* 11:384-386].

- 15 Alternatively, the polymerase chain reaction (PCR) is another well-known means for detecting small amounts of target nucleic acids. The assay is described in: Mullis *et al.* [*Meth. Enzymol.* (1987) 155: 335-350]; US patents 4,683,195 and 4,683,202. Two "primer" nucleotides hybridize with the target nucleic acids and are used to prime the reaction. The primers can comprise sequence that does not hybridize to the sequence of the amplification target (or its complement) to aid with  
20 duplex stability or, for example, to incorporate a convenient restriction site. Typically, such sequence will flank the desired Neisserial sequence.

A thermostable polymerase creates copies of target nucleic acids from the primers using the original target nucleic acids as a template. After a threshold amount of target nucleic acids are generated by the polymerase, they can be detected by more traditional methods, such as Southern  
25 blots. When using the Southern blot method, the labelled probe will hybridize to the Neisserial sequence (or its complement).

- Also, mRNA or cDNA can be detected by traditional blotting techniques described in Sambrook *et al* [*supra*]. mRNA, or cDNA generated from mRNA using a polymerase enzyme, can be purified and separated using gel electrophoresis. The nucleic acids on the gel are then blotted onto a solid  
30 support, such as nitrocellulose. The solid support is exposed to a labelled probe and then washed

to remove any unhybridized probe. Next, the duplexes containing the labeled probe are detected. Typically, the probe is labelled with a radioactive moiety.

## BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1-20 show biochemical data obtained in the Examples, and also sequence analysis, for ORFs 37, 5, 2, 15, 22, 28, 32, 4, 61, 76, 89, 97, 106, 138, 23, 25, 27, 79, 85 and 132. M1 and M2 are molecular weight markers. Arrows indicate the position of the main recombinant product or, in Western blots, the position of the main *N.meningitidis* immunoreactive band. TP indicates *N.meningitidis* total protein extract; OMV indicates *N.meningitidis* outer membrane vesicle preparation. In bactericidal assay results: a diamond (♦) shows preimmune data; a triangle (▲) shows GST control data; a circle (●) shows data with recombinant *N.meningitidis* protein. Computer analyses show a hydrophilicity plot (upper), an antigenic index plot (middle), and an AMPHI analysis (lower). The AMPHI program has been used to predict T-cell epitopes [Gao *et al.* (1989) *J. Immunol.* 143:3007; Roberts *et al.* (1996) *AIDS Res Hum Retrovir* 12:593; Quakyi *et al.* (1992) *Scand J Immunol* suppl.11:9) and is available in the Protean package of DNASTAR, Inc. (1228 South Park Street, Madison, Wisconsin 53715 USA).

## EXAMPLES

The examples describe nucleic acid sequences which have been identified in *N.meningitidis*, along with their putative translation products, and also those of *N.gonorrhoeae*. Not all of the nucleic acid sequences are complete *ie.* they encode less than the full-length wild-type protein.

The examples are generally in the following format:

- a nucleotide sequence which has been identified in *N.meningitidis* (strain B)
- the putative translation product of this sequence
- a computer analysis of the translation product based on database comparisons
- corresponding gene and protein sequences identified in *N.meningitidis* (strain A) and in *N.gonorrhoeae*
- a description of the characteristics of the proteins which indicates that they might be suitably antigenic
- results of biochemical analysis (expression, purification, ELISA, FACS *etc.*)

The examples typically include details of sequence identity between species and strains. Proteins that are similar in sequence are generally similar in both structure and function, and the sequence identity often indicates a common evolutionary origin. Comparison with sequences of proteins of known function is widely used as a guide for the assignment of putative protein function to a new sequence and has proved particularly useful in whole-genome analyses.

Sequence comparisons were performed at NCBI (<http://www.ncbi.nlm.nih.gov>) using the algorithms BLAST, BLAST2, BLASTn, BLASTp, tBLASTn, BLASTx, & tBLASTx [eg. see also Altschul *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25:2289-3402]. Searches were performed against the following databases: non-redundant GenBank+EMBL+DDBJ+PDB sequences and non-redundant GenBank CDS translations+PDB+SwissProt+SPupdate+PIR sequences.

To compare Meningococcal and Gonococcal sequences, the tBLASTx algorithm was used, as implemented at [http://www.genome.ou.edu/gono\\_blast.html](http://www.genome.ou.edu/gono_blast.html). The FASTA algorithm was also used to compare the ORFs (from GCG Wisconsin Package, version 9.0).

Dots within nucleotide sequences (eg. position 495 in SEQ ID 11) represent nucleotides which have been arbitrarily introduced in order to maintain a reading frame. In the same way, double-underlined nucleotides were removed. Lower case letters (eg. position 496 in SEQ ID 11) represent ambiguities which arose during alignment of independent sequencing reactions (some of the nucleotide sequences in the examples are derived from combining the results of two or more experiments).

Nucleotide sequences were scanned in all six reading frames to predict the presence of hydrophobic domains using an algorithm based on the statistical studies of Esposti *et al.* [Critical evaluation of the hydropathy of membrane proteins (1990) *Eur J Biochem* 190:207-219]. These domains represent potential transmembrane regions or hydrophobic leader sequences.

Open reading frames were predicted from fragmented nucleotide sequences using the program ORFFINDER (NCBI).

Underlined amino acid sequences indicate possible transmembrane domains or leader sequences in the ORFs, as predicted by the PSORT algorithm (<http://www.psort.nibb.ac.jp>). Functional domains were also predicted using the MOTIFS program (GCG Wisconsin & PROSITE).



Various tests can be used to assess the *in vivo* immunogenicity of the proteins identified in the examples. For example, the proteins can be expressed recombinantly and used to screen patient sera by immunoblot. A positive reaction between the protein and patient serum indicates that the patient has previously mounted an immune response to the protein in question *ie.* the protein is an immunogen. This method can also be used to identify immunodominant proteins.

The recombinant protein can also be conveniently used to prepare antibodies *eg.* in a mouse. These can be used for direct confirmation that a protein is located on the cell-surface. Labelled antibody (*eg.* fluorescent labelling for FACS) can be incubated with intact bacteria and the presence of label on the bacterial surface confirms the location of the protein.

10 In particular, the following methods (A) to (S) were used to express, purify and biochemically characterise the proteins of the invention:

#### A) Chromosomal DNA preparation

*N.meningitidis* strain 2996 was grown to exponential phase in 100ml of GC medium, harvested by centrifugation, and resuspended in 5ml buffer (20% Sucrose, 50mM Tris-HCl, 50mM EDTA, pH8).  
15 After 10 minutes incubation on ice, the bacteria were lysed by adding 10ml lysis solution (50mM NaCl, 1% Na-Sarkosyl, 50µg/ml Proteinase K), and the suspension was incubated at 37°C for 2 hours. Two phenol extractions (equilibrated to pH 8) and one  $\text{ChCl}_3$ /isoamylalcohol (24:1) extraction were performed. DNA was precipitated by addition of 0.3M sodium acetate and 2 volumes ethanol, and was collected by centrifugation. The pellet was washed once with 70%  
20 ethanol and redissolved in 4ml buffer (10mM Tris-HCl, 1mM EDTA, pH 8). The DNA concentration was measured by reading the OD at 260 nm.

#### B) Oligonucleotide design

Synthetic oligonucleotide primers were designed on the basis of the coding sequence of each ORF, using (a) the meningococcus B sequence when available, or (b) the gonococcus/meningococcus A  
25 sequence, adapted to the codon preference usage of meningococcus as necessary. Any predicted signal peptides were omitted, by deducing the 5'-end amplification primer sequence immediately downstream from the predicted leader sequence.

For most ORFs, the 5' primers included two restriction enzyme recognition sites (*Bam*HI-*Nde*I, *Bam*HI-*Nhe*I, or *Eco*RI-*Nhe*I, depending on the gene's own restriction pattern); the 3' primers included

a *XhoI* restriction site. This procedure was established in order to direct the cloning of each amplification product (corresponding to each ORF) into two different expression systems: pGEX-KG (using either *BamHI-XhoI* or *EcoRI-XhoI*), and pET21b+ (using either *NdeI-XhoI* or *NheI-XhoI*).

5'-end primer tail: CGCGGATCCCATATG (*BamHI-NdeI*)

5 CGCGGATCCGCTAGC (*BamHI-NheI*)

CCGGAATTCTAGCTAGC (*EcoRI-NheI*)

3'-end primer tail: CCCGCTCGAG (*XhoI*)

For ORFs 5, 15, 17, 19, 20, 22, 27, 28, 65 & 89, two different amplifications were performed to clone each ORF in the two expression systems. Two different 5' primers were used for each ORF; the same 3' *XhoI* primer was used as before:

5'-end primer tail: GGAATTCCATATGGCCATGG (*NdeI*)

5'-end primer tail: CGGGATCC (*BamHI*)

ORF 76 was cloned in the pTRC expression vector and expressed as an amino-terminus His-tag fusion. In this particular case, the predicted signal peptide was included in the final product. *NheI*-  
15 *BamHI* restriction sites were incorporated using primers:

5'-end primer tail: GATCAGCTAGCCATATG (*NheI*)

3'-end primer tail: CGGGATCC (*BamHI*)

As well as containing the restriction enzyme recognition sequences, the primers included nucleotides which hybridized to the sequence to be amplified. The number of hybridizing  
20 nucleotides depended on the melting temperature of the whole primer, and was determined for each primer using the formulae:

$$T_m = 4 (G+C) + 2 (A+T) \quad (\text{tail excluded})$$

$$T_m = 64.9 + 0.41 (\% \text{ GC}) - 600/N \quad (\text{whole primer})$$

The average melting temperature of the selected oligos were 65-70°C for the whole oligo and  
25 50-55°C for the hybridising region alone.

Table I (page 487) shows the forward and reverse primers used for each amplification. In certain cases, it will be noted that the sequence of the primer does not exactly match the sequence in the ORF. When initial amplifications were performed, the complete 5' and/or 3' sequence was not

known for some meningococcal ORFs, although the corresponding sequences had been identified in gonococcus. For amplification, the gonococcal sequences could thus be used as the basis for primer design, altered to take account of codon preference. In particular, the following codons were changed: ATA→ATT; TCG→TCT; CAG→CAA; AAG→AAA; GAG→GAA; CGA→CGC; CGG→CGC; GGG→GGC. Italicised nucleotides in Table I indicate such a change. It will be appreciated that, once the complete sequence has been identified, this approach is generally no longer necessary.

Oligos were synthesized by a Perkin Elmer 394 DNA/RNA Synthesizer, eluted from the columns in 2ml NH<sub>4</sub>OH, and deprotected by 5 hours incubation at 56°C. The oligos were precipitated by addition of 0.3M Na-Acetate and 2 volumes ethanol. The samples were then centrifuged and the pellets resuspended in either 100µl or 1ml of water. OD<sub>260</sub> was determined using a Perkin Elmer Lambda Bio spectrophotometer and the concentration was determined and adjusted to 2-10pmol/µl.

### C) Amplification

The standard PCR protocol was as follows: 50-200ng of genomic DNA were used as a template in the presence of 20-40µM of each oligo, 400-800µM dNTPs solution, 1x PCR buffer (including 1.5mM MgCl<sub>2</sub>), 2.5 units *TaqI* DNA polymerase (using Perkin-Elmer AmpliTaq, GIBCO Platinum, Pwo DNA polymerase, or Tahara Shuzo Taq polymerase).

In some cases, PCR was optimised by the addition of 10µl DMSO or 50µl 2M betaine.

After a hot start (adding the polymerase during a preliminary 3 minute incubation of the whole mix at 95°C), each sample underwent a double-step amplification: the first 5 cycles were performed using as the hybridization temperature the one of the oligos excluding the restriction enzymes tail, followed by 30 cycles performed according to the hybridization temperature of the whole length oligos. The cycles were followed by a final 10 minute extension step at 72°C.

The standard cycles were as follows:

	Denaturation	Hybridisation	Elongation
First 5 cycles	30 seconds 95°C	30 seconds 50-55°C	30-60 seconds 72°C
Last 30 cycles	30 seconds	30 seconds	30-60 seconds

	95°C	65-70°C	72°C
--	------	---------	------

The elongation time varied according to the length of the ORF to be amplified.

The amplifications were performed using either a 9600 or a 2400 Perkin Elmer GeneAmp PCR System. To check the results, 1/10 of the amplification volume was loaded onto a 1-1.5% agarose gel and the size of each amplified fragment compared with a DNA molecular weight marker.

- 5 The amplified DNA was either loaded directly on a 1% agarose gel or first precipitated with ethanol and resuspended in a suitable volume to be loaded on a 1% agarose gel. The DNA fragment corresponding to the right size band was then eluted and purified from gel, using the Qiagen Gel Extraction Kit, following the instructions of the manufacturer. The final volume of the DNA fragment was 30µl or 50µl of either water or 10mM Tris, pH 8.5.

#### 10 D) Digestion of PCR fragments

The purified DNA corresponding to the amplified fragment was split into 2 aliquots and double-digested with:

- *NdeI/XhoI* or *NheI/XhoI* for cloning into pET-21b+ and further expression of the protein as a C-terminus His-tag fusion
- 15 – *BamHI/XhoI* or *EcoRI/XhoI* for cloning into pGEX-KG and further expression of the protein as N-terminus GST fusion.
- For ORF 76, *NheI/BamHI* for cloning into pTRC-HisA vector and further expression of the protein as N-terminus His-tag fusion.
- *EcoRI/PstI*, *EcoRI/SalI*, *SalI/PstI* for cloning into pGex-His and further expression of
- 20 the protein as N-terminus His-tag fusion

- Each purified DNA fragment was incubated (37°C for 3 hours to overnight) with 20 units of each restriction enzyme (New England Biolabs) in a either 30 or 40µl final volume in the presence of the appropriate buffer. The digestion product was then purified using the QIAquick PCR purification kit, following the manufacturer's instructions, and eluted in a final volume of 30 or
- 25 50µl of either water or 10mM Tris-HCl, pH 8.5. The final DNA concentration was determined by 1% agarose gel electrophoresis in the presence of titrated molecular weight marker.

**E) Digestion of the cloning vectors (pET22B, pGEX-KG, pTRC-His A, and pGex-His)**

10µg plasmid was double-digested with 50 units of each restriction enzyme in 200µl reaction volume in the presence of appropriate buffer by overnight incubation at 37°C. After loading the whole digestion on a 1% agarose gel, the band corresponding to the digested vector was purified  
5 from the gel using the Qiagen QIAquick Gel Extraction Kit and the DNA was eluted in 50µl of 10mM Tris-HCl, pH 8.5. The DNA concentration was evaluated by measuring OD<sub>260</sub> of the sample, and adjusted to 50µg/µl. 1µl of plasmid was used for each cloning procedure.

The vector pGEX-His is a modified pGEX-2T vector carrying a region encoding six histidine residues upstream to the thrombin cleavage site and containing the multiple cloning site of the  
10 vector pTRC99 (Pharmacia).

**F) Cloning**

The fragments corresponding to each ORF, previously digested and purified, were ligated in both pET22b and pGEX-KG. In a final volume of 20µl, a molar ratio of 3:1 fragment/vector was ligated using 0.5µl of NEB T4 DNA ligase (400 units/µl), in the presence of the buffer supplied by the manufacturer.  
15 The reaction was incubated at room temperature for 3 hours. In some experiments, ligation was performed using the Boehringer "Rapid Ligation Kit", following the manufacturer's instructions.

In order to introduce the recombinant plasmid in a suitable strain, 100µl *E. coli* DH5 competent cells were incubated with the ligase reaction solution for 40 minutes on ice, then at 37°C for 3 minutes, then, after adding 800µl LB broth, again at 37°C for 20 minutes. The cells were then  
20 centrifuged at maximum speed in an Eppendorf microfuge and resuspended in approximately 200µl of the supernatant. The suspension was then plated on LB ampicillin (100mg/ml ).

The screening of the recombinant clones was performed by growing 5 randomly-chosen colonies overnight at 37°C in either 2ml (pGEX or pTC clones) or 5ml (pET clones) LB broth + 100µg/ml ampicillin. The cells were then pelleted and the DNA extracted using the Qiagen QIAprep Spin  
25 Miniprep Kit, following the manufacturer's instructions, to a final volume of 30µl. 5µl of each individual miniprep (approximately 1g ) were digested with either *NdeI/XhoI* or *BamHI/XhoI* and the whole digestion loaded onto a 1-1.5% agarose gel (depending on the expected insert size), in parallel with the molecular weight marker (1Kb DNA Ladder, GIBCO). The screening of the positive clones was made on the base of the correct insert size.

For the cloning of ORFs 110, 111, 113, 115, 119, 122, 125 & 130, the double-digested PCR product was ligated into double-digested vector using *EcoRI-PstI* cloning sites or, for ORFs 115 & 127, *EcoRI-SaII* or, for ORF 122, *SaII-PstI*. After cloning, the recombinant plasmids were introduced in the *E.coli* host W3110. Individual clones were grown overnight at 37°C in L-broth with 50µl/ml ampicillin.

### G) Expression

Each ORF cloned into the expression vector was transformed into the strain suitable for expression of the recombinant protein product. 1µl of each construct was used to transform 30µl of *E.coli* BL21 (pGEX vector), *E.coli* TOP 10 (pTRC vector) or *E.coli* BL21-DE3 (pET vector), as described above. In the case of the pGEX-His vector, the same *E.coli* strain (W3110) was used for initial cloning and expression. Single recombinant colonies were inoculated into 2ml LB+Amp (100µg/ml), incubated at 37°C overnight, then diluted 1:30 in 20ml of LB+Amp (100µg/ml) in 100ml flasks, making sure that the OD<sub>600</sub> ranged between 0.1 and 0.15. The flasks were incubated at 30°C into gyratory water bath shakers until OD indicated exponential growth suitable for induction of expression (0.4-0.8 OD for pET and pTRC vectors; 0.8-1 OD for pGEX and pGEX-His vectors). For the pET, pTRC and pGEX-His vectors, the protein expression was induced by addition of 1mM IPTG, whereas in the case of pGEX system the final concentration of IPTG was 0.2mM. After 3 hours incubation at 30°C, the final concentration of the sample was checked by OD. In order to check expression, 1ml of each sample was removed, centrifuged in a microfuge, the pellet resuspended in PBS, and analysed by 12% SDS-PAGE with Coomassie Blue staining. The whole sample was centrifuged at 6000g and the pellet resuspended in PBS for further use.

### H) GST-fusion proteins large-scale purification.

A single colony was grown overnight at 37°C on LB+Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture in a water bath shaker and grown overnight. Bacteria were diluted 1:30 into 600ml of fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD<sub>550</sub> 0.8-1. Protein expression was induced with 0.2mM IPTG followed by three hours incubation. The culture was centrifuged at 8000rpm at 4°C. The supernatant was discarded and the bacterial pellet was resuspended in 7.5ml cold PBS. The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and centrifuged again.

The supernatant was collected and mixed with 150µl Glutathione-Sepharose 4B resin (Pharmacia)

(previously washed with PBS) and incubated at room temperature for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml cold PBS for 10 minutes, resuspended in 1ml cold PBS, and loaded on a disposable column. The resin was washed twice with 2ml cold PBS until the flow-through reached OD<sub>280</sub> of 0.02-0.06. The GST-fusion protein was eluted by addition of 700µl cold Glutathione elution buffer (10mM reduced glutathione, 50mM Tris-HCl) and fractions collected until the OD<sub>280</sub> was 0.1. 21µl of each fraction were loaded on a 12% SDS gel using either Biorad SDS-PAGE Molecular weight standard broad range (M1) (200, 116.25, 97.4, 66.2, 45, 31, 21.5, 14.4, 6.5 kDa) or Amersham Rainbow Marker (M2) (220, 66, 46, 30, 21.5, 14.3 kDa) as standards. As the MW of GST is 26kDa, this value must be added to the MW of each GST-fusion protein.

#### I) His-fusion solubility analysis (ORFs 111-129)

To analyse the solubility of the His-fusion expression products, pellets of 3ml cultures were resuspended in buffer M1 [500µl PBS pH 7.2]. 25µl lysozyme (10mg/ml) was added and the bacteria were incubated for 15 min at 4°C. The pellets were sonicated for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and then separated again into pellet and supernatant by a centrifugation step. The supernatant was collected and the pellet was resuspended in buffer M2 [8M urea, 0.5M NaCl, 20mM imidazole and 0.1M NaH<sub>2</sub>PO<sub>4</sub>] and incubated for 3 to 4 hours at 4°C. After centrifugation, the supernatant was collected and the pellet was resuspended in buffer M3 [6M guanidinium-HCl, 0.5M NaCl, 20mM imidazole and 0.1M NaH<sub>2</sub>PO<sub>4</sub>] overnight at 4°C. The supernatants from all steps were analysed by SDS-PAGE.

The proteins expressed from ORFs 113, 119 and 120 were found to be soluble in PBS, whereas ORFs 111, 122, 126 and 129 need urea and ORFs 125 and 127 need guanidinium-HCl for their solubilization.

#### J) His-fusion large-scale purification.

A single colony was grown overnight at 37°C on a LB + Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture and incubated overnight in a water bath shaker. Bacteria were diluted 1:30 into 600ml fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD<sub>550</sub> 0.6-0.8. Protein expression was induced by addition of 1mM IPTG and the culture further incubated for three hours. The culture was centrifuged at 8000rpm at 4°C, the supernatant was discarded and the bacterial pellet was resuspended in 7.5ml of either (i) cold

buffer A (300mM NaCl, 50mM phosphate buffer, 10mM imidazole, pH 8) for soluble proteins or (ii) buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 8.8) for insoluble proteins.

The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed two times and centrifuged again.

- 5 For insoluble proteins, the supernatant was stored at -20°C, while the pellets were resuspended in 2ml buffer C (6M guanidine hydrochloride, 100mM phosphate buffer, 10mM Tris-HCl, pH 7.5) and treated in a homogenizer for 10 cycles. The product was centrifuged at 13000rpm for 40 minutes.

- Supernatants were collected and mixed with 150µl Ni<sup>2+</sup>-resin (Pharmacia) (previously washed with either buffer A or buffer B, as appropriate) and incubated at room temperature with gentle agitation  
10 for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml buffer A or B for 10 minutes, resuspended in 1ml buffer A or B and loaded on a disposable column. The resin was washed at either (i) 4°C with 2ml cold buffer A or (ii) room temperature with 2ml buffer B, until the flow-through reached OD<sub>280</sub> of 0.02-0.06.

- The resin was washed with either (i) 2ml cold 20mM imidazole buffer (300mM NaCl, 50mM  
15 phosphate buffer, 20mM imidazole, pH 8) or (ii) buffer D (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 6.3) until the flow-through reached the O.D<sub>280</sub> of 0.02-0.06. The His-fusion protein was eluted by addition of 700µl of either (i) cold elution buffer A (300mM NaCl, 50mM phosphate buffer, 250mM imidazole, pH 8) or (ii) elution buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 4.5) and fractions collected until the O.D<sub>280</sub> was 0.1. 21µl of each  
20 fraction were loaded on a 12% SDS gel.

#### K) His-fusion proteins renaturation

- 10% glycerol was added to the denatured proteins. The proteins were then diluted to 20µg/ml using dialysis buffer I (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, 2M urea, pH 8.8) and dialysed against the same buffer at 4°C for 12-  
25 14 hours. The protein was further dialysed against dialysis buffer II (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, pH 8.8) for 12-14 hours at 4°C. Protein concentration was evaluated using the formula:

$$\text{Protein (mg/ml)} = (1.55 \times \text{OD}_{280}) - (0.76 \times \text{OD}_{260})$$



**L) His-fusion large-scale purification (ORFs 111-129)**

500ml of bacterial cultures were induced and the fusion proteins were obtained soluble in buffer M1, M2 or M3 using the procedure described above. The crude extract of the bacteria was loaded onto a Ni-NTA superflow column (Quiagen) equilibrated with buffer M1, M2 or M3 depending on the solubilization buffer of the fusion proteins. Unbound material was eluted by washing the column with the same buffer. The specific protein was eluted with the corresponding buffer containing 500mM imidazole and dialysed against the corresponding buffer without imidazole. After each run the columns were sanitized by washing with at least two column volumes of 0.5 M sodium hydroxide and reequilibrated before the next use.

**10 M) Mice immunisations**

20µg of each purified protein were used to immunise mice intraperitoneally. In the case of ORFs 2, 4, 15, 22, 27, 28, 37, 76, 89 and 97, Balb-C mice were immunised with Al(OH)<sub>3</sub> as adjuvant on days 1, 21 and 42, and immune response was monitored in samples taken on day 56. For ORFs 44, 106 and 132, CD1 mice were immunised using the same protocol. For ORFs 25 and 40, CD1 mice were immunised using Freund's adjuvant, rather than Al(OH)<sub>3</sub>, and the same immunisation protocol was used, except that the immune response was measured on day 42, rather than 56. Similarly, for ORFs 23, 32, 38 and 79, CD1 mice were immunised with Freund's adjuvant, but the immune response was measured on day 49.

**N) ELISA assay (sera analysis)**

20 The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 7ml of Mueller-Hinton Broth (Difco) containing 0.25% Glucose. Bacterial growth was monitored every 30 minutes by following OD<sub>620</sub>. The bacteria were let to grow until the OD reached the value of 0.3-0.4. The culture was centrifuged for 10 minutes at 10000rpm. The supernatant was discarded and bacteria were washed once with PBS, resuspended in PBS containing 0.025% formaldehyde, and incubated for 2 hours at room temperature and then overnight at 4°C with stirring. 100µl bacterial cells were added to each well of a 96 well Greiner plate and incubated overnight at 4°C. The wells were then washed three times with PBT washing buffer (0.1% Tween-20 in PBS). 200µl of saturation buffer (2.7% Polyvinylpyrrolidone 10 in water) was added to each well and the plates incubated for 2 hours at 37°C. Wells were washed

three times with PBT. 200µl of diluted sera (Dilution buffer: 1% BSA, 0.1% Tween-20, 0.1% NaN<sub>3</sub> in PBS) were added to each well and the plates incubated for 90 minutes at 37°C. Wells were washed three times with PBT. 100µl of HRP-conjugated rabbit anti-mouse (Dako) serum diluted 1:2000 in dilution buffer were added to each well and the plates were incubated for 90 minutes at 37°C. Wells were washed three times with PBT buffer. 100µl of substrate buffer for HRP (25ml of citrate buffer pH5, 10mg of O-phenildiamine and 10µl of H<sub>2</sub>O) were added to each well and the plates were left at room temperature for 20 minutes. 100µl H<sub>2</sub>SO<sub>4</sub> was added to each well and OD<sub>490</sub> was followed. The ELISA was considered positive when OD<sub>490</sub> was 2.5 times the respective pre-immune sera.

10 **O) FACScan bacteria Binding Assay procedure.**

The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 4 tubes containing 8ml each Mueller-Hinton Broth (Difco) containing 0.25% glucose. Bacterial growth was monitored every 30 minutes by following OD<sub>620</sub>. The bacteria were let to grow until the OD reached the value of 0.35-0.5. The culture was centrifuged for 10 minutes at 4000rpm. The supernatant was discarded and the pellet was resuspended in blocking buffer (1% BSA, 0.4% NaN<sub>3</sub>) and centrifuged for 5 minutes at 4000rpm. Cells were resuspended in blocking buffer to reach OD<sub>620</sub> of 0.07. 100µl bacterial cells were added to each well of a Costar 96 well plate. 100µl of diluted (1:200) sera (in blocking buffer) were added to each well and plates incubated for 2 hours at 4°C. Cells were centrifuged for 5 minutes at 4000rpm, the supernatant aspirated and cells washed by addition of 200µl/well of blocking buffer in each well. 100µl of R-Phicoerytrin conjugated F(ab)<sub>2</sub> goat anti-mouse, diluted 1:100, was added to each well and plates incubated for 1 hour at 4°C. Cells were spun down by centrifugation at 4000rpm for 5 minutes and washed by addition of 200µl/well of blocking buffer. The supernatant was aspirated and cells resuspended in 200µl/well of PBS, 0.25% formaldehyde. Samples were transferred to FACScan tubes and read. The condition for FACScan setting were: FL1 on, FL2 and FL3 off; FSC-H threshold:92; FSC PMT Voltage: E 02; SSC PMT: 474; Amp. Gains 7.1; FL-2 PMT: 539; compensation values: 0.

**P) OMV preparations**

Bacteria were grown overnight on 5 GC plates, harvested with a loop and resuspended in 10 ml 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes and the bacteria disrupted by sonication for 10 minutes on ice (50% duty cycle, 50% output). Unbroken cells were removed by centrifugation at 5000g for 10 minutes and the total cell envelope fraction recovered by centrifugation at 50000g at 4°C for 75 minutes. To extract cytoplasmic membrane proteins from the crude outer membranes, the whole fraction was resuspended in 2% sarkosyl (Sigma) and incubated at room temperature for 20 minutes. The suspension was centrifuged at 10000g for 10 minutes to remove aggregates, and the supernatant further ultracentrifuged at 50000g for 75 minutes to pellet the outer membranes. The outer membranes were resuspended in 10mM Tris-HCl, pH8 and the protein concentration measured by the Bio-Rad Protein assay, using BSA as a standard.

**Q) Whole Extracts preparation**

Bacteria were grown overnight on a GC plate, harvested with a loop and resuspended in 1ml of 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes.

**15 R) Western blotting**

Purified proteins (500ng/lane), outer membrane vesicles (5µg) and total cell extracts (25µg) derived from MenB strain 2996 were loaded on 15% SDS-PAGE and transferred to a nitrocellulose membrane. The transfer was performed for 2 hours at 150mA at 4°C, in transferring buffer (0.3 % Tris base, 1.44 % glycine, 20% methanol). The membrane was saturated by overnight incubation at 4°C in saturation buffer (10% skimmed milk, 0.1% Triton X100 in PBS). The membrane was washed twice with washing buffer (3% skimmed milk, 0.1% Triton X100 in PBS) and incubated for 2 hours at 37°C with mice sera diluted 1:200 in washing buffer. The membrane was washed twice and incubated for 90 minutes with a 1:2000 dilution of horseradish peroxidase labelled anti-mouse Ig. The membrane was washed twice with 0.1% Triton X100 in PBS and developed with the Opti-4CN Substrate Kit (Bio-Rad). The reaction was stopped by adding water.

**S) Bactericidal assay**

MC58 strain was grown overnight at 37°C on chocolate agar plates. 5-7 colonies were collected and used to inoculate 7ml Mueller-Hinton broth. The suspension was incubated at 37°C on a nutator and let to grow until OD<sub>620</sub> was 0.5-0.8. The culture was aliquoted into sterile 1.5ml Eppendorf

tubes and centrifuged for 20 minutes at maximum speed in a microfuge. The pellet was washed once in Gey's buffer (Gibco) and resuspended in the same buffer to an OD<sub>620</sub> of 0.5, diluted 1:20000 in Gey's buffer and stored at 25°C.

50µl of Gey's buffer/1% BSA was added to each well of a 96-well tissue culture plate. 25µl of diluted mice sera (1:100 in Gey's buffer/0.2% BSA) were added to each well and the plate incubated at 4°C. 25µl of the previously described bacterial suspension were added to each well. 25µl of either heat-inactivated (56°C waterbath for 30 minutes) or normal baby rabbit complement were added to each well. Immediately after the addition of the baby rabbit complement, 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 0). The 96-well plate was incubated for 1 hour at 37°C with rotation and then 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 1). After overnight incubation the colonies corresponding to time 0 and time 1 hour were counted.

Table II (page 493) gives a summary of the cloning, expression and purification results.

#### Example 1

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 1>:

```

1  ATGAAACAGA CAGTCAA.AT GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
51  GAACCGACCG GTGTGGNCGG ATGACGTATC GGATTTTCGG GAAAACTTGC
101 A.GCGGCAGC ACAGGGAAAT GCAGCAGCCC AATACAATT TGGCGCAATG
151 TAT.TACAAA GGACGCGCGT GCGCCGGGAT GATGCTGAAG CGGTTCAGATG
201 GTATCGGCAG CCGGCGGAAC AGGGGTTAGC CCAAGCCCAA TACAATTTGG
251 GCTGGATGTA TGCCAACGGG CGCGC.GTGC GCCAAGATGA TACCGAAGCG
301 GTCAGATGGT ATCGGCAGGC GGCAGCGCAG GGGGTTGTCC AAGCCCAATA
351 CAATTTGGGC GTGATATATG CCGAAGGACG TGGAGTGC GC CAAGACGATG
401 TCGAAGCGGT CAGATGGTTT CGGCAGGCGG CAGCGCAGGG GGTAGCCCAA
451 GCCCAAACA ATTTGGGCGT GATGTATGCC GAAAGANCGC GCGTGCGCCA
501 AGACCG...

```

This corresponds to the amino acid sequence <SEQ ID 2; ORF37>:

```

1  MKQTVXMLAA ALIALGLNRP VWXDDVSDFR ENLXAAQGN AAAQYNLGAM
51  YXQRTVRVRD DAEAVRWYRQ PAEQGLAQAO YNLGWMYANG RXVRQDDTEA
101 VRWYRQAAAO GVVQAQYNLG VIYAEGRGVR QDDVEAVRWF RQAAAQGVAG
151 AQNNLGVMYA ERXRVQRD...

```

Further work revealed the complete nucleotide sequence <SEQ ID 3>:

```

1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
51  GAACCGAGCG GTGTGGGCGG ATGACGTATC GGATTTTCGG GAAAACTTGC
101 AGGCGGCAGC ACAGGGAAAT GCAGCAGCCC AATACAATT TGGCGCAATG
151 TATTACAAAG GACGCGCGGT GCGCCGGGAT GATGCTGAAG CGGTTCAGATG
201 GTATCGGCAG CCGGCGGAAC AGGGGTTAGC CCAAGCCCAA TACAATTTGG
251 GCTGGATGTA TGCCAACGGG CGCGGCGTGC GCCAAGATGA TACCGAAGCG
301 GTCAGATGGT ATCGGCAGGC GGCAGCGCAG GGGGTTGTCC AAGCCCAATA
351 CAATTTGGGC GTGATATATG CCGAAGGACG TGGAGTGC GC CAAGACGATG
401 TCGAAGCGGT CAGATGGTTT CGGCAGGCGG CAGCGCAGGG GGTAGCCCAA
451 GCCCAAACA ATTTGGGCGT GATGTATGCC GAAAGACGCG GCGTGCGCCA
501 AGACCGCGCC CTTGCACAAG AATGGTTTGG CAAGGCTTGT CAAAACGGAG
551 ACCAAGACGG CTGCGACAAT GACCAACGCC TGAAGCGGG TTATTGA

```

This corresponds to the amino acid sequence <SEQ ID 4; ORF37-1>:

```

1  MKQTVKWLAA ALIALGLNRA VWADDVSDFR ENLQAAAQGN AAAQYNLGAM
51 YYKGRGVRRD DAEAVRWYRQ AAEQGLAQAA YNLGWMYANG RGVRRQDDTEA
101 VRWYRQAAAQ GVVQAQYNLG VIYAEGRGVR QDDVEAVRWF RQAAAQGVAAQ
5  151 AQNNLGVMYA ERRGVRRDRA LAQEWFGKAC QNGDQDGCND DQRLKAGY*

```

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 5>:

```

1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
51 GAACCAAGCG GTGTGGGCGG ATGACGTATC GGATTTTCGG GAAAACTTGC
101 AGGCGGCAGC ACAGGGAAAT GCAGCAGCCC AAAACAATT GGGCGTGATG
10  151 TATGCCGAAA GACGCGGCGT GCGCCAAGAC CGCGCCCTTG CACAAGAATG
201 GCTTGGCAAG GCTTGTCAAA ACGGATACCA AGACAGCTGC GACAATGACC
251 AACGCCTGAA AGCGGGTTAT TGA

```

This encodes a protein having amino acid sequence <SEQ ID 6; ORF37a>:

```

1  MKQTVKWLAA ALIALGLNQA VWADDVSDFR ENLQAAAQGN AAAQNNLGVM
15  51 YAERRGVRRD RALAQEWLGK ACQNGYQDSC DNDQRLKAGY *

```

The originally-identified partial strain B sequence (ORF37) shows 68.0% identity over a 75aa overlap with ORF37a:

```

20  orf37.pep  MKQTVXMLAAALIALGLNRPVWXDDVSDFRNLXAAAQGNAAAQYNLGAMYXQRTVRVRD
      orf37a   MKQTVKWLAAALIALGLNQAQVWADDVSDFRNLQAAAQGNAAAQNNLGVMYAERRGVRRD
      10      20      30      40      50      60
25  orf37.pep  DAEAVRWYRQPAEQGLAQAYNLGWMYANGRXVRQDDTEAVRWYRQAAAQGVVQAQYNLG
      orf37a   RALAQEWLGKACQNGYQDSCDNDQRLKAGYX
      70      80      90

```

Further work identified the corresponding gene in *N.gonorrhoeae* <SEQ ID 7>:

```

1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
51 GAACCAAGCG GTGTGGGCGG GTGACGTATC GGATTTTCGG GAAAACTTGC
101 AGGCGGCAGC ACAGGGAAAT GCAGCAGCCC AATTCATTT GGGCGTGATG
35  151 TATGAAATG GACAAGGAGT TCGTCAAGAT TATGTACAGG CAGTGCAGTG
201 GTATCGCAAG GCTTCAGAAC AAGGGGATGC CCAAGCCCAA TACAATTTGG
251 GCTTGATGTA TTACGATGGA CGCGGCGTGC GCCAAGACCT TCGGCTCGCT
301 CAACATGGC TTGGCAAGGC TTGTCAAAAC GGAGACCAA ACAGCTGCGA
351 CAATGACCAA CGCCTGAAGG CGGGTTATTA A

```

This encodes a protein having amino acid sequence <SEQ ID 8; ORF37ng>:

```

40  1  MKQTVKWLAA ALIALGLNQA VWAGDVSDFR ENLQAAEQGN AAAQFNLGVM
51  YENGQGVRRD YVQAVQWYRK ASEQGDAAQ YNLGLMYDGR RGVRRDLALA
101 QQWLKGACQN GDQNSCDNDQ RLKAGY*

```

The originally-identified partial strain B sequence (ORF37) shows 64.9% identity over a 111aa overlap with ORF37ng:

```

45  orf37.pep  MKQTVXMLAAALIALGLNRPVWXDDVSDFRNLXAAAQGNAAAQYNLGAMYXQRTVRVRD  60
      orf37ng  MKQTVKWLAAALIALGLNQAQVWAGDVSDFRNLQAAEQGNAAAQFNLGVMYENGQGVRRD  60
50  orf37.pep  DAEAVRWYRQPAEQGLAQAYNLGWMYANGRXVRQDDTEAVRWYRQAAAQGVVQAQYNLG  120
      orf37ng  YVQAVQWYRKASEQGDAAQAYNLGLMYDGRGVRRDLALAQQWLKGACQNGDQNSCDNDQ  120
      orf37.pep  VIYAEGRGVRQDDVEAVRWFRQAAAQGVAAQNNLGVMYAERXVRQD  168
55  orf37ng    RLKAGY  126

```

The complete strain B sequence (ORF37-1) and ORF37ng show 51.5% identity in 198 aa overlap:

		10	20	30	40	50	60
	orf37-1.pep	MKQTVKWLAAALIALGLNRAVWADDVSD	FRENLQAAQGNAAQYNLGAMYKGRGVR	RD			
5	orf37ng	MKQTVKWLAAALIALGLNQAVWAGDVSD	FRENLQAAEQGNAAQFNLGVMYENGQGV	RQD			
		10	20	30	40	50	60
		70	80	90	100	110	120
10	orf37-1.pep	DAEAVRWYRQAAEQGLAQYNLGWMYANG	RGVRRQDDTEAVRWYRQAAQGVVQAQYN	LG			
	orf37ng	YVQAVQWYRKASEQDQAQYNLGLMYDGR	GVRQD				
		70	80	90			
		130	140	150	160	170	180
15	orf37-1.pep	VIYAEGRGVRRQDDVEAVRWFRQAAQGV	AQAQNNLGVMYAERRGVRRQDRALAEW	FGKAC			
	orf37ng					LALAQQLGKAC	
						100	
20		190	199				
	orf37-1.pep	QNGDQDGCNDQRLKAGYX					
	orf37ng	QNGDQNSCDNDQRLKAGYX					
		110	120				

- 25 Computer analysis of these amino acid sequences indicates a putative leader sequence, and it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF37-1 (11kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 1A shows the results of affinity purification of the GST-fusion protein, and Figure 1B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result), FACS analysis (Figure 1C), and a bactericidal assay (Figure 1D). These experiments confirm that ORF37-1 is a surface-exposed protein, and that it is a useful immunogen.

- 35 Figure 1E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF37-1.

## Example 2

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 9>:

40 TCCGGCGA CATCGGCGGT TTGAAGGTCA ATGCCCCCGT CAAATCCGCA  
GGCGTATTGG TCGGGCGCGT CGGCGCTATC GGACTTGACC CGAAATCCTA  
TCAGGCGAGG GTGCGCCTCG ATTTGGACGG CAAGTATCAG TTCAGCAGCG  
ACGTTTCCGC GCAAATCCTG ACTTCsGGAC TTTTGGGCGA GCAGTACATC  
GGGCTGCAGC AGGGCGGCGA CACGGAAAAC CTTGCTGCCG GCGACACCAT  
CTCCGTAACC AGTTCTGCAA TGGTTCTGGA AAACCTTATC GGCAAATTCA  
45 TGACGAGTTT TGCCGAGAAA AATGCCGACG GCGCAATGC GGAAAAAGCC  
GCCGAATAA

This corresponds to the amino acid sequence <SEQ ID 10>:

1 FGDIGGLKVN APVKSAGVLV GRVGAIGLDP KSYQARVRLD LDGKYQFSSD  
51 VSAQILTSGL LGEQYIGLQQ GGDENLAAG DTISVTSSAM VLENLIGKFM

101 TSFAEKNADG GNAEKAAE\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a hypothetical *H.influenzae* protein (ybrd.haein; accession number p45029)

SEQ ID 9 and ybrd.haein show 48.4% aa identity in 122 aa overlap:

```

5      20      30      40      50      60      70
yrbd.h LGIGALVFLGLRVANVQGF AETKSYT VTATFDNIGGLKVRAPLKIGGVVIGRVSAITLDE
N.m      FGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
              10      20      30

10     80      90      100     110     120     130
yrbd.h KSYLPKVSIAINQEYNEIPENSSLSIKTSGLLGEQYIALTMGFDDGDTAMLKNGSQIQDT
N.m      KSYQARVRDLDDGKY-QFSSDVSAQILTSGLLGEQYIGLQQG---GDTENLAAGDTISVT
              40      50      60      70      80

15     140     150     160
yrbd.h TSAMVLEDLIGQFL--YGSKKSDGNEKSESTEQ
N.m      SSAMVLENLIGKFMTSFAEKNADGGNAEKAAEX
              90      100     110     120

```

Homology with a predicted ORF from *N.gonorrhoeae*

SEQ ID 9 shows 99.2% identity over a 118aa overlap with a predicted ORF from *N. gonorrhoeae*:

```

25     20      30      40      50      60      70
yrbd      GAAAVAFLAFRVAGGAAFGGSDKTYAVYADFGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
N.m      FGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
              10      20      30

30     80      90      100     110     120     130
yrbd      KSYQARVRDLDDGKYQFSSDVSAQILTSGLLGEQYIGLQQGGDTENLAAGDTISVTSSAM
N.m      KSYQARVRDLDDGKYQFSSDVSAQILTSGLLGEQYIGLQQGGDTENLAAGDTISVTSSAM
              40      50      60      70      80      90

35     140     150     160
yrbd      VLENLIGKFMTSFAEKNAE G GNAEKAAEX
N.m      VLENLIGKFMTSFAEKNADGGNAEKAAEX
              100     110     120

```

The complete yrbd *H.influenzae* sequence has a leader sequence and it is expected that the full-length homologous *N.meningitidis* protein will also have one. This suggests that it is either a membrane protein, a secreted protein, or a surface protein and that the protein, or one of its epitopes, could be a useful antigen for vaccines or diagnostics.

**Example 3**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 11>:

```

50      1  ..ATTTTGATAT ACCTCATCCG CAAGAATCTA GGTTCGCCCG TCTTCTTCTT
      51  TCAGGAACGC CCCGGAAGG ACGGAAAACC TTTTAAAATG GTCAAATTCC
      101  GTTCCATGCG CGACGGCTTG TATTCAGACG GCATTCGCT GCGCGACGGA
      151  GAACGCCTGA CACCGTTCGG CAAAAAAGT CGTGCCGcCA GTwTGGACGA
      201  ACTGCCTGAA TTATGGAATA TCTTAAAAGG CGAGATGAGC CTGGTCGGCC
      251  CCGCCCCGCT GCTGATGCAA TATCTGCCGC TGTACGACAA CTTCCAAAC
      301  CGCGGCCACG AAATGAAACC CGGCATTACC GGCTGGGCGC AGGTCAACGG

```

5  
10

```

351 GCGCAACGCG CTTTCGTGGG ACGAAAAATT CGCCTGCGAT GTTGGGTATA
401 TCGACCACTT CAGCCTGTGC CTCGACATCA AAATCCTACT GCTGACGGTT
451 AAAAAAGTAT TAATCAAGGA AGGGATTTC GCACAGGGCG AACAA.aCCAT
501 GCCCCCTTTC ACAGGAAAAC GCAAACTCGC CGTCGTGCGT GCGGGCGGAC
551 ACGGAAAAGT CGTTGCCGAC CTTGCCGCG CACTCGGCCG GTACAGGGAA
601 ATCGTTTTTC TGGACGACCG CGCACAGGC AGCGTCAACG GCTTTTCCGT
651 CATCGGCACG ACGCTGTGTC TTGAAAACAG TTTATCGCCC GAACAATACG
701 ACGTCGCGGT CGCCGTGCGC AACAAACGCA TCCGCCGCCA AATCGCCGAA
751 AAAGCCGCG CGCTCGGCTT CGCCCTGCCC GTACTGGTTC ATCCGGACGC
801 GACCGTCTCG CCTTCTGCAA CAGTCGGACA AGGCAGCGTC GTTATGGCGA
851 AAGCGGTGCG.

```

This corresponds to the amino acid sequence <SEQ ID 12; ORF3>:

15

```

1  .ILIYLIRKNL GSPVFFFQER PGKDGKPFKM VKFRSMRDGL YSDGIPLPDG
51  ERLTFPGKKL RAASXDELPE LWNILKGEMS LVGPRPLLMQ YLPLYDNFQN
101 RRHEMKPGIT GWAQVNGRNA LSWDEKFACD VWYIDHFSLC LDIKILLTLV
151 KKVLIKEGIS AQEXTMPFF TGRKRLAVVG AGGHGKVVD LAAALGRYRE
201 IVFLDDRAQG SVNGFSVIGT TLLLENSLSP EQYDVAVAVG NNRIRRQIAE
251 KAAALGFALP VLVHPDATVS PSATVGQGSV VMAKAV..

```

Further sequence analysis revealed the complete nucleotide sequence <SEQ ID 13>:

20  
25  
30  
35  
40

```

1  ATGAGTAAAT TCTTCAAACG CCTGTTTGAC ATTGTTGCCT CCGCCTCGGG
51  ACTGATTTTC CTCTCGCCAG TATTTTGTAT TTTGATATAC CTCATCCGCA
101 AGAATCTAGG TTCGCCCCTC TTCTTCTTTC AGGAACGCCC CGGAAAGGAC
151 GGAAAACCTT TTAATAATGGT CAAATTCCTG TCCATGCGCG ACGCGCTTGA
201 TTCAGACGGC ATTCCGCTGC CCGACGGAGA ACGCCTGACA CCGTTCGGCA
251 AAAAAGTTCG TGCCGCCAGT TTGGACGAAC TGCTGAATT ATGGAATATC
301 TTAAGAGGCG AGATGAGCCT GGTCCGCCCC CGCCCGCTGC TGATGCAATA
351 TCTGCCGCTG TACGACAAC TCCAAAACCG CCGCCACGAA ATGAAACCCG
401 GCATTACCGG CTGGGCGCAG GTCAACGGGC GCAACGCGCT TTCGTGGGAC
451 GAAAAATTCG CCTGCGATGT TTGGTATATC GACCACTTCA GCCTGTGCCT
501 CGACATCAAA ATCCTACTGC TGACGGTTAA AAAAGTATTA ATCAAGGAAG
551 GGATTTCCGC ACAGGCGGAA GCCACCATGC CCCCTTTCAC AGGAAAACGC
601 AAACCTCGCCG TCGTCGGTGC GGGCGGACAC GGAAAAGTCG TTGCCGACCT
651 TGCCGCCGCA CTCGGCCGGT ACAGGGAAAT CGTTTTTCTG GACGACCGCG
701 CACAAGGCAG CGTCAACGGC TTTTCCGTCA TCGGCACGAC GCTGCTGCTT
751 GAAAAACAGT TATCGCCCGA ACAATACGAC GTCGCCGTCG CCGTCGGCAA
801 CAACCGCATC CGCCGCCAAA TCGCCGAAAA AGCCGCCGCG CTCGGCTTCG
851 CCCTGCCCGT TCTGGTTCAT CCGGACGCGA CCGTCTCGCC TTCTGCAACA
901 GTCGGACAAG GCAGCGTCGT TATGGCGAAA GCCGTCGTAC AGGCAGGCAG
951 CGTATTGAAA GACGGCGTGA TTGTGAACAC TGCCGCCACC GTCGATCAGC
1001 ACTGCCTGCT TAACGCTTTC GTCCACATCA GCCCAGGCGC GCACCTGTGC
1051 GGCAACACGC ATATCGGCGA AGAAAGCTGG ATAGGCACGG GCGCGTGCAG
1101 CCGCCAGCAG ATCCGTATCG GCAGCCGCGC AACCATTGGA GCGGGCGCAG
1151 TCGTCGTACG CGACGTTTCA GACGGCATGA CCGTCGCGGG CAATCCGGCA
1201 AAGCCGCTGC CGCGCAAAAA CCCCAGAGACC TCGACAGCAT AA

```

45 This corresponds to the amino acid sequence <SEQ ID 14; ORF3-1>:

50

```

1  MSKFFKRLFD IVASASGLIF LSPVFLILY LIRKNLGSFV FFFQERPGKD
51  GKPFKMVKFR SMRDALDS DG IPLPDGERLT PFGKKLRAAS LDELPELWNI
101 LKGEMSLVGP RPLLMQYLPL YDNFQNRHE MKPGITGWAQ VNGRNALSWD
151 EKFCADVWYI DHFSLCLDIK ILLLTVKKVL IKEGISAQGE ATMPFFTGRK
201 KLAVVGAGGH GKVVDLAAA LGRYREIVFL DDRAQGSVNG FSVIGTLLLL
251 ENSLSPEQYD VAVAVGNNRI RRQIAEKAAA LGFALPVLVH PDATVSPSAT
301 VGQGSVVMK AVVQAGSVLK DGVIVNTAAT VDHDCLLNAF VHISFGAHL
351 GNTHIGEESE IGTGACSRQQ IRIGSRATIG AGAVVVRDVS DGMTVAGNPA
401 KPLPRKNPET STA*

```

55 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF3 shows 93.0% identity over a 286aa overlap with an ORF (ORF3a) from strain A of *N. meningitidis*:



-65-

orf3.pep  
 orf3a MSKFFKRLFDIVASASGLIFLSPVFLILILIYLRKNLGSVPVFFQERPGKDGKPFKMVKFR  
 10 20 30 40 50 60

5  
 orf3.pep SMRDGLYSDGIPLPDGERLTPFGKKLRAASKDELPELWNILKGEMSLVGPRPLLMQYLPL  
 40 50 60 70 80 90  
 orf3a SMHDA LDSGILLPDGERLTPFGKKLRAASLDELPELWNVLKGDMSLVGPRPLLMQYLPL  
 70 80 90 100 110 120

10  
 orf3.pep YDNFQNR RHMKPGITGWAQVNGRNALSWDEKFA CDVWYIDHFSLC LDIKILLTVKKVL  
 100 110 120 130 140 150  
 orf3a YDNFQNR RHMKPGITGWAQVNGRNALSWDERFACDIWYIDHFSLC LDIKILLTVKKVL  
 130 140 150 160 170 180

15  
 orf3.pep IKEGISAQGE XTMPPTGKRKLAVVGAGGHGKV VADLAAALGRYREIVFLDDRAQGSVNG  
 160 170 180 190 200 210  
 orf3a IKEGISAQGEATMPPTGKRKLAVVGAGGHGKVVAELAAALGTYGEIVFLDDRQGSVNG  
 190 200 210 220 230 240

20  
 orf3.pep FSVIGTTLLEN SLSPEQYDVAVVGNNRIR RQIAEKAAALGFALPVLVHPDATVSPSAT  
 220 230 240 250 260 270  
 orf3a FSVIGTTLLEN SLSPEQFDIAVVGNNRIR RQIAEKAAALGFALPVLVHPDSTVSPSAT  
 250 260 270 280 290 300

25  
 orf3.pep VGQGSVVMAKAV  
 280  
 orf3a VGQGSVVMAKAVVQADSVLKDGIVNTAATVDHDCLLDAFVHISPGAHLSGNTRIGEEESW  
 310 320 330 340 350 360

35 The complete length ORF3a nucleotide sequence <SEQ ID 15> is:

1 ATGAGTAAAT TCTTCAAACG CCTGTTTGAC ATTGTTGCCT CGCCTCGGG  
 51 ACTGATTTC CTCTCGCCAG TATTTTGTAT TTTGATATAC CTCATCCGCA  
 101 AGAATCTGGG TTCGCCGTC TTCTTCTTTC AGGAACGCC CGGAAAGGAC  
 151 GGAAAACCTT TTAATATGGT CAAATCCGT TCCATGCACG ACGCGCTTGA  
 40 201 TTCAGACGGC ATTCTGCTGC CCGACGAGA ACGCTGACA CCGTTCGGCA  
 251 AAAAAGTGC TGCCGCCAGT TTGGACGAAC TGCCGGAAC GTGGAACGTC  
 301 CTCAAAGGCG ACATGAGCCT GGTCCGCCCC CGCCCGCTGC TGATGCAATA  
 351 TCTGCCGCTG TACGACAACT TCCAAAACCG CCGCCACGAA ATGAAACCGG  
 401 GCATTACCGG CTGGGCGCAG GTCAACGGGC GCAACGCGCT TTCGTGGGAC  
 45 451 GAACGCTTCG CATGCGACAT CTGGTATATC GACCACTTCA GCCTGTGCCT  
 501 CGACATCAAA ATCTACTGCT TGACGGTTAA AAAAGTATTA ATCAAAGAAG  
 551 GGATTTCCGC ACAGGCGGAA GCCACCATGC CCCCTTTCAC AGGAAAACGC  
 601 AAACCTGCGG TCGTCGGTGC GGGCGGACAC GGCAAAGTCG TTGCCGAGCT  
 651 TGCCGCCGCA CTCGGCACAT ACGGCGAAAT CGTTTTCTG GACGACCGCG  
 50 701 TCCAAGGCAG CGTCAACGGC TTCCCGTCA TCGGCACGAC GCTGCTGCTT  
 751 GAAAACAGTT TATCGCCCGA ACAATTCGAC ATCGCCGTCG CCGTCGGCAA  
 801 CAACCGCATC CGCCGCCAAA TCGCCGAAAA AGCCGCCGCG CTCGGCTTCG  
 851 CCCTGCCCGT CCTGATTCAT CCGGACTCGA CCGTCTCGCC TTCTGCAACA  
 901 GTCGGACAAG GCGGCGTCGT TATGGCGAAA GCCGTCGTAC AGGCTGACAG  
 55 951 CGTATTGAAA GACGGCGTAA TTGTGAACAC TGCCGCCACC GTCGATCAGC  
 1001 ATGTCCTGCT TGATGCTTTC GTCCACATCA GCCCGGGCGC GCACCTGTCTG  
 1051 GGCAACACGC GTATCGGCGA AGAAAGCTGG ATAGGCACAG GCGCGTGCAG  
 1101 CCGCCAGCAG ATCCGTATCG GCAGCCGCGC AACCATTGGA GCGGCGCAG  
 1151 TCGTCGTGCG CGACGTTTCA GACGGCATGA CCGTCGCGGG CAACCCGGCA  
 60 1201 AAACCATGAG CAGGCAAAAA TACCGAGACC CTGCGGTCGT AA

This is predicted to encode a protein having amino acid sequence <SEQ ID 16>:

1 MSKFFKRLFD IVASASGLIF LSPVFLILY LIRKNLGSVP FFFQERPGKD  
 51 GKPFKMVKFR SMHDA LDSG ILLPDGERLT PFGKKLRAAS LDELPELWNV  
 101 LKGDMSLVGP RPLLMQYLPL YDNFQNR RHE MKPGITGWAQ VNGRNALSWD  
 65 151 ERFACDIWYI DHFSLCLDIK ILLTVKKVL IKEGISAQGE ATMPPTGKR  
 201 KLA VVGAGGH GKVVAELAAA LGTYGEIVL DDRVQGSVNG FSVIGTTL  
 251 ENSLSPEQFD IAVAVGNNRI RRQIAEKAAA LGFALPVLH PDSTVSPSAT

301 VGQGGVVMK AVVQADSVLK DGVI VNTAAT VDHDCLLDAF VHISPGAHLS  
 351 GNTRIGEE SW IGTGACSRQQ IRIGSRATIG AGAVVVRDVS DGMTVAGNPA  
 401 KPLAGKNTET LRS\*

Two transmembrane domains are underlined.

5 ORF3-1 shows 94.6% identity in 410 aa overlap with ORF3a:

		10	20	30	40	50	60
	orf3a.pep	MSKFFKRLFDIVASASGLIFLSPVFLILYLRKKNL	GSPVFFFQERPGKD	GKPFKMKVKFR			
	orf3-1	MSKFFKRLFDIVASASGLIFLSPVFLILYLRKKNL	GSPVFFFQERPGKD	GKPFKMKVKFR			
10		10	20	30	40	50	60
	orf3a.pep	SMHDA	LDSDGILLPDGERLTPFGKKLRAASLDELPELWNVLKGDMSLVGPRPLLMQYLPL				
	orf3-1	SMRDA	LDSDGILPLDGERLTPFGKKLRAASLDELPELWNILKGEMSLVGPRPLLMQYLPL				
15		70	80	90	100	110	120
	orf3a.pep	YDNFQNR	RHEMKPGITGWAQVNGRNALSWDERFACDIWYIDHFSCLCDIKILLTVKKVL				
	orf3-1	YDNFQNR	RHEMKPGITGWAQVNGRNALSWDEKFA	CDVWYIDHFSCLCDIKILLTVKKVL			
20		130	140	150	160	170	180
	orf3a.pep	IKEGISAQGEATMP	PFTGKRKLAVVGAGGHGKVVAAELAAALGTYGEIVFLDDRQGSVNG				
	orf3-1	IKEGISAQGEATMP	PFTGKRKLAVVGAGGHGKVVADLAAALGRYREIVFLDDRQGSVNG				
25		190	200	210	220	230	240
	orf3a.pep	FPVIGTTLL	ENSLSP	EQFDIAVAVGNNRIRRQIAEKAAALGFALPVLHPDSTVSPSAT			
	orf3-1	FSVIGTTLL	ENSLSP	EQYDVAVAVGNNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT			
30		250	260	270	280	290	300
	orf3a.pep	VGQGGVVMK	AVVQADSVLKDGVI	VNTAATVDHDCLLDAFVHISPGAHLSGNTRIGEE	SW		
	orf3-1	VGQGSVVMK	AVVQAGSVLKDGVI	VNTAATVDHDCLLNAFVHISPGAHLSGNTHIGEE	SW		
35		310	320	330	340	350	360
	orf3a.pep	IGTGACSRQQ	IRIGSRATIGAGAVVVRDVS	DGMTVAGNPAKPLAGKNTETLRSX			
	orf3-1	IGTGACSRQQ	IRIGSRATIGAGAVVVRDVS	DGMTVAGNPAKPLPRKNPETSTAX			
40		370	380	390	400	410	
	orf3a.pep	IGTGACSRQQ	IRIGSRATIGAGAVVVRDVS	DGMTVAGNPAKPLAGKNTETLRSX			
	orf3-1	IGTGACSRQQ	IRIGSRATIGAGAVVVRDVS	DGMTVAGNPAKPLPRKNPETSTAX			
45		370	380	390	400	410	

Homology with hypothetical protein encoded by yvfc gene (accession Z71928) of *B. subtilis*

ORF3 and YVFC proteins show 55% aa identity in 170 aa overlap (BLASTp):

50	ORF3	3	IYLIRKKNL	GSPVFFFQERPGKD	GKPFKMKVKFR	SMRDGLYSDGILPDGERLTPFGKKLRA	62
	yvfc	27	IYVRLKIG	SPVFFKQVRPGLHGKPF	TLTKFTMTDERDSKGNLLPDEVRLTKTGRLIRK	86	
	ORF3	63	ASXDELPELWNILKGEMSLVGPRPLLMQYL	PLYDNFQNR	RHEMKPGITGWAQVNGRNALS	122	
55	yvfc	87	LSIDELPQLLNVLKGDLSLVGPRPLMDYL	PLYTEKQARRHEVKPGITGWAQINGRNAIS	146		
	ORF3	123	WDEKFA	CDVWYIDHFSCLDXXXXXXXXXXXXXXXXX	EGISAQGE	XTMP	FTG
	yvfc	147	WEKKFEL	DVWYVDNWSFFLDL	KILCLTVRKVLVSEGIQ	TNHVTAER	FTG
60							

Homology with a predicted ORF from *N.gonorrhoeae*

ORF3 shows 86.3% identity over a 286aa overlap with a predicted ORF (ORF3.ng) from *N. gonorrhoeae*:

5	orf3	ILIIYLIRKNLGSPVFFFQERPGKDGPVKMKVFR	34
	orf3ng	MSKAVKRLFDIIASASGLIVLSPVFLVLIYLIRKNGSPVFFIRERPGKDGPVKMKVFR	60
10	orf3	SMRDGLYSDGIPLPDGERLTPFGKKLRAASXDELPELWNILKGEMSLVGPRPLMQYLPL	94
	orf3ng	SMRDALDSDGIPLDSERLTDFGKKLRATSLDELPELWNVLKGEMSLVGPRPLMQYLPL	120
15	orf3	YDNFQNRHHEMKPGITGWAQVNGRNALSWDEKFACDVWYIDHFSCLDIKILLTVKKVL	154
	orf3ng	YNKFQNRHHEMKPGITGWAQVNGRNALSWDEKFSVDVWYTDNFSFWLDMKILFLTVKKVL	180
20	orf3	IKEGISAQGEATMPPTGKRKLAVVGAGGHGKVVAADLAAALGRYREIVFLDDRAQGSVNG	214
	orf3ng	IKEGISAQGEATMPPFAGNRKLAVIGAGGHGKVVAELAAALGTYGEIVFLDDRTQGSVNG	240
25	orf3	FSVIGTTLLENSLSPEQYDVAVAVGNNRIRRQIAEKAAALGFALPVLVHPDATVSPSAI	274
	orf3ng	FPVIGTTLLENSLSPEQFDITVAVGNNRIRRQITENAAALGFKLPVLIHPDATVSPSAI	300
30	orf3	VGQGSVVMKAV	286
	orf3ng	IGQGSVVMKAVVQAGSVLKDGIVNTAATVDHCLLDADFVHISPGAHLGNTTRIGEEER	360

The complete length ORF3ng nucleotide sequence <SEQ ID 17> is:

	1	ATGAGTAAAG	CCGTCAAACG	CCTGTTCGAC	ATCATCGCAT	CCGCATCGGG
	51	GCTGATTGTC	CTGTGCGCCG	TGTTTTTGGT	TTTAATATAC	CTCATCCGCA
30	101	AAACTTAGG	TTCGCCCCGTC	TTCTTCattc	GGGAACGCCc	cgGAAAGGAc
	151	ggaaaaaCTT	TTAAATAGGT	CAAATTCCTG	TCCAtgcgcg	acgcgcttGA
	201	TTCAAGCGGC	ATTCCGCTGC	CCGATAGCGA	ACGCCTGACC	GATTTCGGGA
	251	AAAAATTACG	CGCCACCAGT	TTGGACGAAC	TTCTGAATT	ATGGAATGTC
	301	CTCAAAGCGC	AGATGAGCCT	GGTCGGCCCC	CGCCCGCTTT	TGATGCAGTA
35	351	TCTGCCGCTT	TACAACAAAT	TTCAAACCG	CGCCACGAA	ATGAAACCGG
	401	GCATTACCGG	CTGGGCGCAG	GTCAACGGGC	GCAACGCGCT	TTCTGTTGGC
	451	GAAAGTTTCT	CCTGCGATGT	TTGGTACACC	GACAATTTCA	GCTTTTGGCT
	501	GGATATGAAA	ATCCTGTTTC	TGACAGTCAA	AAAAGTCTTG	ATTAAAGAAG
	551	GCATTTCGGC	GCAAGGGGAA	GCCACCATGC	CCCCTTTCGC	GGGGAATCGC
40	601	AAACTCGCCG	TTATCGGCCG	GGGCGGACAC	GGCAAAGTCG	TTGCCGAGCT
	651	TGCCGCGGCA	CTCGGCACAT	ACGGCGAAAT	CGTTTTTCTG	GACGACCGCA
	701	CCCAAGGCAG	CGTCAACGGC	TTCCCGCTCA	TCGGCAGGCT	GCTGCTGCTT
	751	GAAACAGT	TATCGCCCGA	ACAATTCGAC	ATCACCGTCG	CCGTCGGCAA
	801	CAACCGCATC	CGCCGCCAAA	TCACCGAAAA	CGCCGCCCGG	CTCGGCTTCA
45	851	AACTGCCCGT	TCTGATTTCAT	CCCGACGCGA	CCGTCTCGCC	TTCTGCAATA
	901	ATCGGACAAG	GCAGCGTCGT	AATGGCGAAA	GCCGTCGTAC	AGGCCGCGAG
	951	CGTATTGAAA	GACGGCGTGA	TTGTGAACAC	TGCCGCCACC	GTGATCAGC
	1001	ACTGCCTGCT	TGACGCTTTC	GtccaCATCA	GCCCGGGCGC	GCACCTGTCG
	1051	GGCAACACGC	GTATCGGCGA	AGAAAGCCGG	ATAGGCACGG	GCGCGTGCAG
50	1101	CCGCCAGCAG	ACAACCGTGC	GCAGCGGGGT	TACCgcccGT	GCAGGGgcGG
	1151	TTATCGTATG	CGACATCCCG	GACGGCATGA	CCGTGCGGGG	CAACCCGGCA
	1201	AAGCCCTTA	CGGGCAAAAA	CCCCAAGACC	GGGACGGCAT	AA

This encodes a protein having amino acid sequence <SEQ ID 18>:

	1	MSKAVKRLFD	IIASASGLIV	LSPVFLVLIY	LIRKNLGSPV	FFIRERPGKD
55	51	GKPFKMKVFR	SMRDALDSG	IPLPDSERLT	DFGKKLRATS	LDELPELWNV
	101	LKGEMSLVGP	RPLLMQYLPL	YNKFQNRHHE	MKPGITGWAQ	VNGRNALSWD
	151	EKFSCDVWYT	DNFSFWLDMK	ILFLTVKKVL	IKEGISAQGE	ATMPPFAGNR
	201	KLAVIGAGGH	GKVVAELAAA	LGTYGEIVFL	DDRTQGSVNG	FPVIGTTLIL
	251	ENSLSPEQFD	ITVAVGNNRI	RRQITENAAA	LGFKLPVLIH	PDATVSPSAI
	301	IGQGSVMAK	AVVQAGSVLK	DGVIVNTAAT	VDHCLLDADF	VHISPGAHL
60	351	GNTTRIGEEER	IGTGACSRQQ	TTVGSGVTAG	AGAVIVCDIP	DGMTVAGNPA
	401	KPLTGKNPKT	GTA*			

This protein shows 86.9% identity in 413 aa overlap with ORF3-1:

		10	20	30	40	50	60
	orf3-1.pep	MSKFFKRLFDIVASASGLIFLSPVFLIYLIYLRKKNLGSPPVFFQERPGKDGKPFKMKVFR					
5	orf3ng	MSKAVKRLFDIIASASGLIVLSPVFLVLIYLIYLRKKNLGSPPVFFIRERPGKDGKPFKMKVFR					
		10	20	30	40	50	60
		70	80	90	100	110	120
10	orf3-1.pep	SMRDALDSGDIPLPDGERLTDFGKKLRATSLDELPELWNVLKGEMSLVGPRLMLQYLYPL					
	orf3ng	SMRDALDSGDIPLPDSERLTDFGKKLRATSLDELPELWNVLKGEMSLVGPRLMLQYLYPL					
		70	80	90	100	110	120
		130	140	150	160	170	180
15	orf3-1.pep	YDNFQNRHMKPGITGWAQVNGRNALSWDEKFCADVWYIDHFSCLDIKILLTVKKVL					
	orf3ng	YNKFQNRHMKPGITGWAQVNGRNALSWDEKFCADVWYIDHFSCLDIKILLTVKKVL					
		130	140	150	160	170	180
		190	200	210	220	230	240
20	orf3-1.pep	IKEGISAQGEATMPFFTGKRKLAVVGAGGHGKVADLAAALGRYREIVFLDDRAQGSVNG					
	orf3ng	IKEGISAQGEATMPFFAGNRKLAVIGAGGHGKVVAELAAALGTGEIVFLDDRTQGSVNG					
		190	200	210	220	230	240
25		250	260	270	280	290	300
	orf3-1.pep	FSVIGTTLLENLSLSPQYDVAVAVGNNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT					
	orf3ng	FPVIGTTLLENLSLSPQFDITVAVGNNRIRRQITENAAALGFKLPVLVHPDATVSPSAI					
30		250	260	270	280	290	300
		310	320	330	340	350	360
	orf3-1.pep	VGQGSVMAKAVVQAGSVLKDGVIVNTAATVDHDCLLNAFVHISPGAHLGSGNTHIGEEESW					
35	orf3ng	IGQGSVMAKAVVQAGSVLKDGVIVNTAATVDHDCLLDAFVHISPGAHLGSGNTRIGEEESR					
		310	320	330	340	350	360
		370	380	390	400	410	
40	orf3-1.pep	IGTGACSRQQIRIGSRATIGAGAVVVRDVS DGMTVAGNPAKPLPRKNPETSTAX					
	orf3ng	IGTGACSRQQTTVGSGVTAGAGAVIVCDIPDGMTVAGNPAKPLTGKNPKTGTX					
		370	380	390	400	410	

In addition, ORF3ng shows significant homology with a hypothetical protein from *B. subtilis*:

45	gnl PID e238668 (271928) hypothetical protein [Bacillus subtilis]
	>gi 1945702 gnl PID e313004 (294043) hypothetical protein [Bacillus subtilis]
	>gi 2635938 gnl PID e1186113 (299121) similar to capsular polysaccharide biosynthesis [Bacillus subtilis] Length = 202
	Score = 235 bits (594), Expect = 3e-61
	Identities = 114/195 (58%), Positives = 142/195 (72%)
50	Query: 5 VKRLFDIIASASGLIVLSPVFLVLIYLIYLRKKNLGSPPVFFIRERPGKDGKPFKMKVFRSMRD 64
	+KRLFD+ A+ L S + L I ++R +GSPVEF + RPG GKPF + KFR+M D
	Sbjct: 3 LKRLFDLTAAIFLLCCTSVIILEFTIAVRLKIGSPVFFKQVRPGLHGKPFITYKFTMTD 62
55	Query: 65 ALDSGDIPLPDSERLTDFGKKLRATSLDELPELWNVLKGEMSLVGPRLMLQYLYPLYNKF 124
	DS G LPD RLT G+ +R S+DELP+L NVLKG++SLVGPRLML YLPLY +
	Sbjct: 63 ERDSKGNLLPDEVRLTKTGRILIRKLSIDELPQLLNVLKGDLSLVGPRLMLDYLYPTEK 122
60	Query: 125 QNRHMKPGITGWAQVNGRNALSWDEKFCADVWYIDHFSCLDIKILLTVKKVLKEG 184
	Q RRHE+KPGITGWAQ+NGRNA+SW++KF DVWY DN+SF+LD+KIL LTV+KVL+ EG
	Sbjct: 123 QARRHEVKPGITGWAQINGRNAISWEKKFELDVWYVDNWSFFLDLKLCLTVRKVLVSEG 182
	Query: 185 ISAQGEATMPFFAGN 199
	I T F G+
65	Sbjct: 183 IQQTNHVTAERTGS 197

The hypothetical product of *yvf* gene shows similarity to EXOY of *R.meliloti*, an exopolysaccharide production protein. Based on this and on the two predicted transmembrane regions in the homologous *N.gonorrhoeae* sequence, it is predicted that these proteins, or their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### 5 Example 4

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 19>:

```

1  ..AACCATATGG CGATTGTCAT CGACGAATAC GCGGGCACAT CCGGCTTGGT
51  CACCTTTGAA GACATCATCG AGCAAATCGT CCGCGAAATC GAAGACGAGT
101 TTAGCAAGA CGATAGCGCC GACAAATATCC ATGCCGTTTC TTCAGACACG
151 TGGCGCATCC ATGCAGCTAC CGAAATCGAA GACATCAACA CCTTCTTCGG
201 CACGGAATAC AGCATCGAAG AAGCCGACAC CATT.GGCGG CCTGGTCATT
251 CAAGAGTTGG GACATCTGCC CGTGCGCGGC GAAAAAGTCC TTATCGCGCG
301 TTTGCAGTTC ACCGTCGCAC GCGCCGACAA CCGCCGCCTG CATACGCTGA
351 TGGCGACCCG CGTGAAGTAA GC..... ACCGC CGTTTCTGCA
15 401 CAGTTTAG

```

This corresponds to amino acid sequence <SEQ ID 20; ORF5>:

```

1  ..NHMAIVIDEY GGTSGLVTFE DIIEQIVGEI EDEFDEDDSA DNIHAVSSDT
51  WRIHAATEIE DINTFFGTEY SIEEADTIXR PGHSRVGTS RARRKSPYRR
101 FAVHRRTRRQ PPPAYADGDP REVS....XR RCTV*

```

20 Further sequence analysis revealed the complete DNA sequence to be <SEQ ID 21>:

```

1  ATGGACGGCG CACAACCGAA AACGAATTTT TTTGAACGCC TGATTGCCCG
51  ACTCGCCGCG GAACCCGATT CCGCCGAAGA CGTATTAAAC CTGCTTCGGC
101 AGGCGCACGA GCAGGAAGTT TTTGATGCGG ATACGCTTTT AAGATTGGAA
151 AAAGTCCTCG ATTTTCCGA TTTGGAAGTG CGCGACGCGA TGATTACGCG
25 201 CAGCCGTATG AACGTTTTAA AAGAAAACGA CAGCATCGAG CGCATCACCG
251 CCTACGTTAT CGATACCGCC CATTGCGGCT TCCCCGTCAT CGGCGAAGAC
301 AAAGACGAAG TTTTGGGCAT TTTGCACGCC AAAGACCTGC TCAAATATAT
351 GTTTAACCCC GAGCAGTTCC ACCTCAAATC CATTCTCCGC CCCGCCGTCT
401 TCGTCCCCGA AGGCAAATCG CTGACCGCCC TTTTAAAAGA GTTCCGCGAA
30 451 CAGCGCAACC ATATGGCGAT TGTATCGAC GAATACGGCG GCACATCCGG
501 CTTGGTCACC TTTGAAGACA TCATCGAGCA AATCGTCGGC GAAATCGAAG
551 ACGAGTTTGA CGAAGACGAT AGCGCCGACA ATATCCATGC CGTTTCTTCC
601 GAACGCTGGC GCATCCATGC AGCTACCGAA ATCGAAGACA TCAACACCTT
651 CTTGCGCACG GAATACAGCA GCGAAGAAGC CGACACCATT CGGCCTGGTC
35 701 ATTCAGAGT TGGGACATCT GCCCGTGCAG GCGGAAAAG TCCTTATCGG
751 CGGTTTGAG TACACGTCG CACGCGCCGA CAACCGCCGC CTGCATACGC
801 TGATGGGAC CCGGTGAAG TAAGCACGCG CGTTTCTGCA CAGTTTAGGA
851 TGACGGTACG GCGGTTTTCT GTTTCATCC GCCCATCCG CCAAACATAA

```

This corresponds to amino acid sequence <SEQ ID 22; ORF5-1>:

```

40 1  MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLLRLE
51  KVLDFSLELV RDAMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
101 KDEVLGILHA KDLLKYMFPN EQFHLKSILR PAVFVPEGKS LTALLKEFRE
151 QRNHMAIVID EYGGTSGLVT FEDIIEQIVG EIEDEFDEDD SADNIHAVSS
201 ERWRIHAATE IEDINTFFGT EYSSEADTI RPHGSRVGTS ARARRKSPYR
45 251 RFAVHRRTRR QPPPAYADGD PREVSTAVSA QFRMTVRAFS VSIRPIRQT*

```

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 23 >:

```

1  ATGGACGGCG CACAACCGAA AACAAATTTT TTNNAACGCC TGATTGCCCG
51  ACTCGCCGCG GAACCCGATT CCGCCGAAGA CGTATTGACC CTGTTGCGCC
101 AAGCGCACGA ACAGGAAGTA TTTGATGCGG ATACGCTTTT AAGATTGGAA
151 AAAGTCCTCG ATTTTCTGA TTTGGAAGTG CGCGACGCGA TGATTACGCG
50 201 CAGCCGTATG AACGTTTTAA AAGAAAACGA CAGCATCGAA CGCATCACCG
251 CCTACGTTAT CGATACCGCC CATTGCGGCT TCCCCGTCAT CGGTGAAGAC
301 AAAGACGAAG TTTTGGGTAT TTTGCACGCC AAAGACCTGC TCAAATATAT
351 GTTCAACCCC GAGCAGTTCC ACCTCAAATC GATATTGCGC CCTGCGGTCT

```

401 TCGTCCCCGA AGGCAAATCG CTGACCGCCC TTTTAAAAGA GTTCCGCGAA  
 451 CAGCGCAACC ATATGGCAAT CGTCATCGAC GAATACGGCG GCACGTCGGG  
 501 TTTGGTAACT TTTGAAGACA TCATCGAGCA AATCGTCGGC GACATCGAAG  
 551 ATGAGTTTGA CGAAGACGAA AGCGCGGACA ACATCCACGC CGTTTCCGCC  
 601 GAACGCTGGC GCATCCACGC GGCTACCGAA ATCGAAGACA TCAACGCCTT  
 651 TTTCCGACAG GAATACAGCA GCGAAGAAGC CGACACCATC GCGCGCCNTG  
 701 GTCATTACAG AATTGGNACA CCTGCCCCGTG CGCGGCGAAA AAGTCNTTAT  
 751 CGGCGNNTTG CANTTCACNG TCGCCNGCGC NGACAACCGC CGCCTGCATA  
 801 CGCTGATGGC GACCCGCGTG AAGTAAGCTC CGCCGTTTCT GTACAGTTTA  
 10 851 GGATGACGGT ACGGGCGTTT TCTGTTTCAA TCCGCCCAT CCGCCANACA  
 901 TAA

This encodes a protein having amino acid sequence <SEQ ID 24; ORF5a>:

1 MDGAQPKTNF XXRLIARLAR EPDSAEDVLT LLROAHEQEV FDADTLRLLE  
 51 KVLDFSDLEV RDAMITSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED  
 15 101 KDEVLGILHA KDLKYMFP EQFHLKSILR PAVFVPEGKS LTALLKEFRE  
 151 QRNHMAIVID EYGGTSGLVT FEDIIEQIVG DIEDEFDEDE SADNIHAVSA  
 201 ERWRIHAATE IEDINAFGT EYSSEADTI GGXGHSIGT PARARRKSYX  
 251 RRAXHXRXR XQPPAYADG DPREVSSAVS VQFRMTVRAF SVSIRPIRXT  
 301 \*

20 The originally-identified partial strain B sequence (ORF5) shows 54.7% identity over a 124aa overlap with ORF5a:

					10	20	30
	orf5.pep				NHMAIVIDEYGGTSGLVT	FEDIIEQIVGEI	
25	orf5a	FHLKSILRPAV	FVPEGKSLTALLKEFREQRNHMAIV	IDEYGGTSGLVT	FEDIIEQIVGDI		
		130	140	150	160	170	180
		40	50	60	70	80	90
	orf5.pep	EDEFDEDSADNIHAVSSDTWRIHAATEIEDINTFFGTEYSIEEADTIXRPGHSRVGTSA					
30							
	orf5a	EDEFDEDESADNIHAVSAERWRIHAATEIEDINAFGT EYSSEADTIGGXGHSIGTPA					
		190	200	210	220	230	240
		100	110	120	130		
35	orf5.pep	RARRKSPYRRFAVHRRTTRQPPAYADGDPREVSSXXXXRRFCTV					
	orf5a	RARRKSXYRRAXHXRXRQPPAYADGDPREVSSAVSVQFRMTVRAFSVSIRPIRXTX					
		250	260	270	280	290	300

The complete strain B sequence (ORF5-1) and ORF5a show 92.7% identity in 300 aa overlap:

					10	20	30	40	50	60
40	orf5a.pep	MDGAQPKTNFXXRLIARLAREPDSAEDVLTLLROAHEQEVFDADTLRLLEKVLDFSDLEV								
	orf5-1	MDGAQPKTNFFERLIARLAREPDSAEDVLNLLROAHEQEVFDADTLRLLEKVLDFSDLEV								
		10	20	30	40	50	60			
45		70	80	90	100	110	120			
	orf5a.pep	RDAMITSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYMFP								
	orf5-1	RDAMITSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYMFP								
50		70	80	90	100	110	120			
		130	140	150	160	170	180			
	orf5a.pep	EQFHLKSILRPAV	FVPEGKSLTALLKEFREQRNHMAIV	IDEYGGTSGLVT	FEDIIEQIVG					
55	orf5-1	EQFHLKSILRPAV	FVPEGKSLTALLKEFREQRNHMAIV	IDEYGGTSGLVT	FEDIIEQIVG					
		130	140	150	160	170	180			
		190	200	210	220	230	240			
	orf5a.pep	DIEDEFDEDESADNIHAVSAERWRIHAATEIEDINAFGT EYSSEADTIGGXGHSIGT								
60	orf5-1	EIEDEFDEDSADNIHAVSSERWRIHAATEIEDINTFFGTEYSSEADTIRP-GHSRVGT								
		190	200	210	220	230				
		250	260	270	280	290	300			

```

orf5a.pep  PARARRKSYRRXAXHXRXRQPPPAYADGDPREVSSAVSVQFRMTVRAFSVSIRPIRXT
          ||||| ||| | | | | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf5-1     SARARRKSPYRRFAVHRRTRRQPPPAYADGDPREVSTAVSAQFRMTVRAFSVSIRPIRQT
          240      250      260      270      280      290

```

- 5 Further work identified the a partial DNA sequence in *N.gonorrhoeae* <SEQ ID 25> which encodes a protein having amino acid sequence <SEQ ID 26; ORF5ng>:

```

1  MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLTRLE
51 KVLDFAELEV RDAMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
101 KDEVLGILHA KDLLKYMFPN EQFHLKSVLR PAVFVPEGKS LTALLKEFRE
10  151 QRNHMAIVID EYGGTSGSLVT FEDIIEQIVG DIEDEFDEDE SADDIHSVSA
201 ERWRIHAATE IEDINAFFGT EYGSEEADTI RRLGHSGIGT PARARRKSPY
251 RRFVHRRPR RQPPAHADG DPREVSRACP HRRFCTV*

```

Further analysis revealed the complete gonococcal nucleotide sequence <SEQ ID 27> to be:

```

1  ATGGACGGCG CACAACCGAA AACAAATTTT TTTGAACGCC TGATTGCCCCG
15  51 ACTCGCCCGC GAACCCGATT CCGCCGAAGA CGTATTAAAC CTGCTTCGGC
101 AGGCGCACGA ACAGGAAGTT TTTGATGCCG ACACACTGAC CCGGCTGGAA
151 AAAGTATTGG ACTTTGCCGA GCTGGAAGTG CGCGATGCGA TGATTACGGC
201 CAGCCGCATG AACGTATTGA AAGAAAACGA CAGCATCGAA CGCATCACCG
251 CCTACGTCAT CGATACCGCC CATTGCGGCT TCCCGTCAT CGGCGAAGAC
20  301 AAAGACGAAG TTTTGGGCAT TTTGCACGCC AAAGACCTGC TCAAATATAT
351 GTTCAACCCC GAGCAGTTCC ACCTGAAATC CGTCTTGCGC CTGCCGTTT
401 TCGTGCCCGA AGGCAAATCT TTGACCGCCC TTTTAAAGA GTTCCGCGAA
451 CAGCGCAACC ATATGGCAAT CGTCATCGAC GAATACGGCG GCACGTCGGG
501 TTTGGTCACC TTTGAAGACA TCATCGAGCA AATCGTCGGT GACATCGAAG
25  551 ACGAGTTTGA CGAAGACGAA AGCGccgacg acatCCACTC cgTTTccgCC
601 GAACGCTGGC GCATCCacgc ggctaCCGAA ATCGAAGaca TCAACGCCTT
651 TTTCGGTACG GAatacggca gcgaagaagc cgacaccatc cggcggctTG
701 GTCATTcAGG AATTGGGACA CCTGCCCGTG CGCGGCGAAA AAGTCCTTat
25  751 cggcgGTTG Cagttcaccg tCGCCGCGC CGACAACCGC GCCTGCACA
30  801 CGTGATGGC GACCCGCGTG AAGTAAGCAG AGCCTGCCcg AccgcccgttT
851 CTGCacAGTt TAGGatgACG gtaCGGTCGT TTTCTGTTTC AATCCGCCCC
901 ATCCGCCAAA CATAA

```

This encodes a protein having amino acid sequence <SEQ ID 28; ORF5ng-1>:

```

1  MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLTRLE
35  51 KVLDFAELEV RDAMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
101 KDEVLGILHA KDLLKYMFPN EQFHLKSVLR PAVFVPEGKS LTALLKEFRE
151 QRNHMAIVID EYGGTSGSLVT FEDIIEQIVG DIEDEFDEDE SADDIHSVSA
201 ERWRIHAATE IEDINAFFGT EYGSEEADTI RRLGHSGIGT PARARRKSPY
251 RRFVHRRPR RQPPAHADG DPREVSRACP TAVSAQFRMT VRSFSVSIRP
40  301 IRQT*

```

The originally-identified partial strain B sequence (ORF5) shows 83.1% identity over a 135aa overlap with the partial gonococcal sequence (ORF5ng):

```

orf5              NHMAIVIDEYGGTSGSLVTFEDIIEQIVGEI 30
45  orf5ng        FHLKSVLRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGSLVTFEDIIEQIVGDI 182
orf5              EDEFDEDDSDNIHAVSSDTWRIHAATEIEDINTFFGTEYSIEEADTIXRPGHSRVGTSA 90
50  orf5ng        EDEFDEDESADDIHSVSAERWRIHAATEIEDINAFFGTEYGSSEADTIRRLGHSGIGTTPA 242
orf5              RARRKSPYRRFAVHRRTRRQPPPAYADGDPREVSVX---RRFCTV 131
orf5ng            RARRKSPYRRFAVHRRPRRQPPPAHADGDPREVSRACPHRRFCTV 287

```

The complete strain B and gonococcal sequences (ORF5-1 & ORF5ng-1) show 92.4% identity in

55 304 aa overlap:

```

          10      20      30      40      50      60
orf5ng-1.pep MDGAQPKTNFFERLIARLAREPDSAEDVLNLLRQAHEQEVFDADTLTRLEKVLDFAELEV

```

15

20

25

30

Computer analysis of these amino acid sequences indicates a putative leader sequence, and identified the following homologies:

Homology with hemolysin homolog TlyC (accession U32716) of *H.influenzae*

ORF5 and TlyC proteins show 58% aa identity in 77 aa overlap (BLASTp).

```

40      ORF5      2      HMAIVIDEYGGTSGLVTFEDIIEQIVGEIEDEFEDEDDSDADNIHAVSSDTWRIHAATEIED 61
      TlyC      166    HMAIV+DE+G  SGLVT EDI+EQIVG+IEDEFEDE++ AD I  +S  T+  +  A  T+I+D
      ORF5      62      INTFFGTEYSIEEADTI 78
      TlyC      225    N  F  T++  EE  DTI
      TlyC      225    FNAQFNTDFDDEEVDTI 241

```

45 ORF5ng-1 also shows significant homology with TlyC:

SCORES        Init1:   301 Initn:   419 Opt:   668  
Smith-Waterman score: 668;     45.9% identity in 242 aa overlap

			10	20	30	40	50
50	orf5ng-1.pep		MDGAQPKTNFFERLIARLAR-EPD	SAEDVLNLLRQAHEQEVFDADTL	TRLEK		
			:   :   :   : : : :   :				
	tlyc_haein	MNDEQQNSNQSENTKKPFFQSLFGR	FFQ	GELKNREELVEVIRDSEQN	DLIDONTRE	MIEG	
		10	20	30	40	50	60
55		60	70	80	90	100	109
	orf5ng-1.pep	VLDFAELEV	RDAMITSRMNV	LKENDS	IERITAYVIDT	AHSRFPVIGE--DKDEV	LGI
		::   :         :	: : : : : : : : : : : : : : : :   : : :				
	tlyc_haein	VMEIAELRVR	DIMIPRSQI	IFIEDQQDLNTCLNTI	ESHSRFPVIAD	DRDNIVG	ILH
		70	80	90	100	110	120
60		110	120	130	140	150	160
	orf5ng-1.pep	AKDLLKYM	F-NPEQFHLKSVLR	PAVFVPEGKSLTALLKE	FRQNRHMAIV	IDEYGGT	SGL
		: :       :   :   :   :   :   : :   :   :   :					
	tlyc_haein	AKDLLKFL	REDAEVFDLSSLLRP	VVIVPESKRVD	RMLKDFRSERF	HMAIVVDE	F



```

130      140      150      160      170      180
170      180      190      200      210      220
5  orf5ng-1.pep VTFEDIEIQIVGDIEDEFDEDESADDIHSVSAERWRIHAATEIEDINAFFGTEYGSEEAD
    ||:||||:|||||||:| | |::| : : :| |:|:| | |::| :|:|
tlyc_haein     VTIEDILEIQIVGDIEDEFDEEEIAD-IRQLSRHTYAVRALTDIDDFNAQFNTDFDDEEVD
              190      200      210      220      230
230      240      250      260      270      280
10 orf5ng-1.pep TIRRLGHSGIG-TPARARRKSPYRRFAVHRRPFRQPPPAHADGDPREVSRACTAVSAQF
    || | : : | | :
tlyc_haein     TIGGLIMQTFGYLPKRGEIILKNLQFKVTSADSRRLIQLRVTVPDEHLAEMNNVDEKSE
              240      250      260      270      280      290

```

# 15 Homology with a hypothetical secreted protein from *E.coli*:

ORF5a shows homology to a hypothetical secreted protein from *E.coli*:

```

20 sp|P77392|YBEX_ECOLI_HYPOTHETICAL_33.3_KD_PROTEIN_IN_CUTE-ASNB_INTERGENIC_REGION
>gi|1778577 (U82598) similar to H. influenzae [Escherichia coli] >gi|1786879
(AE000170) f292; This 292 aa ORF is 23% identical (9 gaps) to 272 residues of an
approx. 440 aa protein YTFL_HAEIN SW: P44717 [Escherichia coli] Length = 292

```

Score = 212 bits (533), Expect = 3e-54

Identities = 112/230 (48%), Positives = 149/230 (64%), Gaps = 3/230 (1%)

```

25 Query: 2  DGAQPKTNFXRLIARLAR-EPDSAEDVLTLLRQAHEQEVFDADTLLRLEKVLDFSDLEV 60
    D   K F L+++L EP + +++L L+R + + ++ D DT LE V+D +D V
Sbjct: 10  DTISNKKGFFSLLLSQLFHGEPKRNDELLALIRDSGQNDLIDEDTRDMLEGVMDIADQRV 69

```

```

30 Query: 61  RDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYM-FN 119
    RD MI RS+M LK N +++ +I++AHSRFPVI EDKD + GIL AKDLL +M +
Sbjct: 70  RDIMIPRSQMITLKRNTLDECLDVIIESAHSRFPVISEDKDHIIEGILMAKDLLPFMRSD 129

```

```

35 Query: 120 PEQFHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIEIQIV 179
    E F + +LR AV VPE K + +LKEFR QR HMAIVIDE+GG SGLVT EDI+E IV
Sbjct: 130 AEAFSMDKVLQRQAVVVPESKVRDMLKEFRSQRHMAIVIDEFGGVSGLVTTIEDILELIV 189

```

Query: 180 GDIEDEFDEDESADNIHAVSAERWRIHAATEIEDINAFFGTEYSSEEADT 229

G+IEDE+DE++ D +S W + A IED N FGT +S EE DT

Sbjct: 190 GEIEDEYDEEDDID-FRQLSRHTWTVRALASIEDFNEAFGTHFSDEEVD 238

40 Based on this analysis, including the amino acid homology to the TlyC hemolysin-homologue from *H. influenzae* (hemolysins are secreted proteins), it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae* are secreted and could thus be useful antigens for vaccines or diagnostics.

45 ORF5-1 (30.7kDa) was cloned in the pGex vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 2A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot analysis (Figure 1B). These experiments confirm that ORF5-1 is a surface-exposed protein, and that it is a useful immunogen.

## Example 5

50 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 29>:

```

1  ATGCGCGGCG GCAGGCCGGA TTCCGTTACC GTGCAGATTA TCGAAGGTTG
51 GCGTTTTTCG CATATGAGGA AAGTCATCGA CGCAACGCCC GACATCGGAC

```

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55

```

101 ACGACACCAA AGGCTGGAGC AATGAAAAAC TGATGGCGGA AGTTGCGCCC
151 GATGCCTTCA GCGGCAATCC TGAAGGGCAG TTTTCCCCG ACAGCTACGA
201 AATCGATGCG GCGGCGAGTG ATTTGCAGAT TTACCAAACC GCCTACAAGG
251 GCGATGCAAC GCCGCTGAA TGAGGGCATG GGAAAGCAGG CAGGACGGGC
301 TGCCTTATAA AAACCCCTTAT GAAATGCTGA TTATGGCGAa CCTGGTCGAA
351 AAGGAAACAG GGCATGAAGC CGAsCsCGAC CATGTcGCTT CCGTCTTCGT
401 CAACCGCCTG AAAATCGGTA TCGCCTGCA AACCGAssCG TCCGTGATTT
451 ACGGCATGGG TCGGCATAC AAGGGCAAAA TCCGTAAAGC CGACCTGCGC
501 CGCGACACGC CGTACAACAC CTACACGCGC GGCGGTCTGC CGCCAACCCC
551 GATTGCGCTG CCC..

```

This corresponds to the amino acid sequence <SEQ ID 30; ORF7>:

15  
20  
25  
30  
35  
40  
45  
50  
55

```

1 MRGGRPDsvT VQIIEGSRFS HMRKVIDATP DIGHDTKWS NEKLMAEVAP
51 DAFSGNPEGQ FFPDSYEIDA GGSdLQIYQT AYKAMQRRLN EAWESRQDGL
101 PYKNPYEMLI MAXLVEKETG HEAXXDHVAS VEVNRLKIGM RLQTXSVIY
151 GMGAAYKGKI RKADLRDTP YNTYTRGGLP PTPIALP..

```

Further sequence analysis revealed the complete DNA sequence <SEQ ID 31>:

20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95

```

1 ATGTTGAGAA AATTGTTGAA ATGGTCTGCC GTTTTTTTGA CCGTGTCCGC
51 AGCCGTTTTT GCCGCGCTGC TTTTGTTC TAAGGATAAC GGCAGGGCAT
101 ACCGAATCAA AATTGCCAAA AACCAGGGTA TTTCGTCTCGT CGGCAGGAAA
151 CTTGCCGAAG ACCGCATCGT GTTCAGCAGG CATGTTTTGA CGGCGCGGC
201 CTACGTTTTG GGTGTGCACA ACAGGCTGCA TACGGGGACG TACAGATTGC
251 CTTCCGAAGT GTCTGCTTGG GATATCTTGC AGAAAATGCG CGGCGGCAGG
301 CCGGATTCCG TTACCGTGCA GATTATCGAA GGTTCTCGCTT TTTTCGCATAT
351 GAGGAAAGTC ATCGACGCAA CGCCCGACAT CGGACACGAC ACCAAAGGCT
401 GGAGCAATGA AAACTGATG GCGGAAGTTG CGCCCGATGC CTTCAGCGGC
451 AATCCTGAAG GGCAGTTTTT CCCCACAGC TACGAAATCG ATGCGGGCGG
501 CAGTGATTG CAGATTTACC AAACCGCCTA CAAGGCGATG CAACGCCGCC
551 TGAATGAGGC ATGGGAAAGC AGGCAGGACG GGCTGCCTTA TAAAAACCTT
601 TATGAAATGC TGATTATGGC GAGCCTGGTC GAAAAGGAAA CAGGGCATGA
651 AGCCGACCGC GACCATGTCT CTTCCGTCTT CGTCAACCGC CTGAAATCG
701 GTATGCGCCT GCAAACCGAC CCGTCCGTGA TTTACGGCAT GGGTGCGGCA
751 TACAAGGGCA AAATCCGTAA AGCCGACCTG CGCCGCGACA CGCCGTACAA
801 CACCTACACG CGCGCGGTC TGCCGCCAAC CCCGATTGCG CTGCCCGGCA
851 AGGCGGCACT CGATGCCGCC GCCCATCCGT CCGGCGAAAA ATACCTGTAT
901 TTCGTGTCCA AAATGGACGG CACGGGCTTG AGCCAGTTCA GCCATGATTT
951 GACCGAACAC AATGCCGCCG TCCGCAAATA TATTTTGAAA AAATAA

```

This corresponds to the amino acid sequence <SEQ ID 32; ORF7-1>:

40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95

```

1 MLRKLLKWSA VELTVSAAVE AALLFVPKDN GRAYRIKIAK NQGISSVGRK
51 LAEDRIVFSR HVLTAAYVL GVHNRLHTGT YRLPSEVSAW DILQKMRGGR
101 PDSVTVQIIE GSRFSHMRKV IDATPDIGHD TKGWSNEKLM AEVAPDAFSG
151 NPEGQFFPDS YEIDAGGSDL QIYQTAYKAM QRRLNEAWES RQDGLPYKNP
201 YEMLIMASLV EKETGHEADR DHVASVFVNR LKIGMRLQTD PSVIYGMGAA
251 YKGKIRKADL RRDTPYNTYT RGGLPPTPIA LPGKAALDAA AHPSGEKYLY
301 FVSKMDGTGL SQFSHDLTEH NAAVRKYILK K*

```

45 Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical protein encoded by yceg gene (accession P44270) of *H.influenzae*

ORF7 and yceg proteins show 44% aa identity in 192 aa overlap:

50  
55

```

ORF7 1 MRGGRPDsvTVQIIEGSRFSHMRKVIDATPDIGHDTKWSNEKLMA-----EVAPDAFSG 55
      + G+ V+ IEG F RK ++ P + K SNE++ A ++ +
yceg 102 LNSGKEVQFNVKWIEGKTFKDWKRDLENAPHLVQTLKDKSNEEIFALLDLDPIDIGQNLK 161

ORF7 56 NPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRLNEAWESRQDGLPYKNPYEMLIMAXLV 115
      N EG +PD+Y +DL++ + + + M++ LN+AW R + LP NPYEMLI+A +V
yceg 162 NVEGWLYPDYNTYTPKSTDLELLKRSARMKKALNKAWNERDEDLPLANPYEMLILASIV 221

ORF7 116 EKETGHEAXXDHVASVFVNRLKIGMRLQTXSVIYGMGAAYKGKIRKADLRDTPYNTYT 175
      EKETG VASVF+NRLK M+LQT +VIYGMG Y G IRK DL TPYNTY
yceg 222 EKETGIANERAKVASVFINRLKAKMKLQTDPTVIYGMGENYNGNIRKKDLETKTPYNTYV 281

```

ORF7 176 RGGLPPTPIALP 187  
GLPPTPIA+P  
yceg 282 IDGLPPTPIAMP 293

The complete length YCEG protein has sequence:

```

5      1 MKKFLIAILL LILILAGVAS FSYYKMTEFV KTPVNVQADE LLTIERGTTS
      51 SKLATLFEQE KLIADGKLLP YLLKLKPELN KIKAGTYSLE NVKTVQDILL
     101 LLNSGKEVQF NVKWIEGKTF KDWKRDLENA PHLVQTLKDK SNEEIFALLD
     151 LPDIGQNLLE KNVEGWLYPD TYNYPKSTD LELLKRSRER MKKALNKAWN
     201 ERDEDLPLAN PYEMLILASI VEKETGIANE RAKVASVFIN RLKAKMKLQT
    10  251 DPTVIYGMGE NYNGNIRKKD LETKTPYNTY VIDGLPPTPI AMPSESSLOA
     301 VANPEKTDYF YFVADGSGGH KFTRNLEHN KAVQEYLRYW RSQKNK
  
```

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF7 shows 95.2% identity over a 187aa overlap with an ORF (ORF7a) from strain A of *N. meningitidis*:

```

                                     10      20      30
    orf7.pep                      MRGGRPDSVTVQIIEGSRFSHMRKVIDATP
                                     |||||
    orf7a                          AAYVLGVHNRHLTGTYRLPSEVSAWDILQKMRGGRPDSVTVQIIEGSRFSHMRKVIDATP
    20      70      80      90     100     110     120

                                     40      50      60      70      80      90
    orf7.pep                      DIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSEIDAGGSDLQIYQTAYKAMQRRLN
    25      |||||
    orf7a                          DIEHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSEIDAGGSDLRIYQIAYKAMQRRLN
                                     130     140     150     160     170     180

                                     100     110     120     130     140     150
    orf7.pep                      EAWESRQDGLPYKNPYEMLIMAXLVEKETGHEAXXDHVASVFNRLKIGMRLQTXSVIY
    30      |||||
    orf7a                          EAWESRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFNRLKIGMRLQTDPSVIY
                                     190     200     210     220     230     240

                                     160     170     180
    orf7.pep                      GMGAAYKGKIRKADLRDTPYNTYTRGGLPPTPIALP
    35      |||||
    orf7a                          GMGAAYKGKIRKADLRDTPYNTYTRGGLPPTPIALPGKAALDAAHPSGEKYLYFVSKM
                                     250     260     270     280     290     300

    orf7a                          DGTGLSQFSHDLTEHNAAVRKYLKKX
                                     310     320     330
  
```

The complete length ORF7a nucleotide sequence <SEQ ID 33> is:

```

      1 ATGTTGAGAA AATTGTTGAA ATGGTCTGCC GTTTTTTTGA CCGTATCGGC
     51 AGCCGTTTTT GCCGCGCTGC TTTTCGTCCC TAAAGACAAC GGCAGGGCAT
    45 101 ACAGGATTAA AATTGCCAAA AACCAGGGTA TTTTCGTGGT CGGCAGGAAA
     151 CTTGCCGAAG ACCGCATCGT GTTCAGCAGG CATGTTTTGA CGGCGGCGGC
     201 CTACGTTTTG GGTGTGCACA ACAGGCTGCA TACGGGGACG TACAGACTGC
     251 CTTCGGAAGT GTCTGCTTGG GATATCTTGC AGAAAATGCG CGGCGGCGAG
     301 CCGGATTCCG TTACCGTGCA GATTATCGAA GGTTTCGCTT TTTTCGCATAT
    50 351 GAGGAAAGTC ATCGACGCAA CGCCCGACAT CGAACACGAC ACCAAAGGCT
     401 GGAGCAATGA AAAACTGATG GCGGAAGTTG CCCCTGATGC CTTACGCGGC
     451 AATCCTGAAG GGCAGTTTTT CCCCACAGC TACGAAATCG ATGCGGGCGG
     501 CAGCGATTTA CGGATTACC AAATCGCCTA CAAGCGGATG CAACGCGGAC
     551 TGAATGAGGC ATGGGAAAGC AGGCAGGACG GGCTGCCTTA TAAAAACCTT
    55 601 TATGAAATGC TGATTATGGC GAGCCTGATC GAAAAGGAAA CAGGGCATGA
     651 AGCCGACCGC GACCATGTCG CTTCCGTCTT CGTCAACCGC CTGAAAATCG
     701 GTATGCGCCT GCAAACCGAC CCGTCCGTGA TTTACGGCAT GGGTGCGGCA
     751 TACAAGGGCA AAATCCGTAA AGCCGACCTG CGCCGCGACA CGCCGTGCAA
     801 CACCTACACG CGCGGCGGTC TGCCGCCAAC CCCGATCGCG CTGCCCGGCA
    60 851 AGGGGGCACT CGATGCCGCC GCCCATCCGT CCGGTGAAAA ATACCTGTAT
     901 TTCGTGTCCA AAATGGACGG TACGGGCTTG AGCCAGTTCA GCCATGATT
     951 GACCGAACAC AACGCCGCCG TTCGCAAATA TATTTTGAAA AAATAA
  
```

This is predicted to encode a protein having amino acid sequence <SEQ ID 34>:

5  
1 MLRKLLKWSA VFLTVSAAVF AALLFVPKDN GRAYRIKIAK NQGISSVGRK  
51 LAEDRIVFSR HVLTAAYVL GVHNLHTGT YRLPSEVSAW DILQKMRGGR  
101 PDSVTVQIIE GSRFSHMRKV IDATPDIEHD TKGWSNEKLM AEVAPDAFSG  
151 NPEGQFFPDS YEIDAGGSDL RIYQIAYKAM QRRLEAWES RQDGLPYKNP  
201 YEMLIMASLI EKETGHEADR DHVASVFVNR LKIGMRLQTD PSVIYMGAA  
251 YKGKIRKADL RRDTPYNTYT RGGLPPTPIA LPGKAALDAA AHPSGEKYLY  
301 FVSKMDGTGL SQFSDLTEH NAAVRKYILK K\*

A leader peptide is underlined.

10 ORF7a and ORF7-1 show 98.8% identity in 331 aa overlap:

15  
20  
25  
30  
35  
40  
45

orf7a.pep	10	20	30	40	50	60
	MLRKLLKWSA	VFLTVSAAVF	AALLFVPKDN	GRAYRIKIAK	NQGISSVGRK	LAEDRIVFSR
orf7-1	10	20	30	40	50	60
	MLRKLLKWSA	VFLTVSAAVF	AALLFVPKDN	GRAYRIKIAK	NQGISSVGRK	LAEDRIVFSR
orf7a.pep	70	80	90	100	110	120
	HVLTAAYVL	GVHNLHTGT	YRLPSEVSAW	DILQKMRGGR	PDSVTVQI	IEGSRFSHMRKV
orf7-1	70	80	90	100	110	120
	HVLTAAYVL	GVHNLHTGT	YRLPSEVSAW	DILQKMRGGR	PDSVTVQI	IEGSRFSHMRKV
orf7a.pep	130	140	150	160	170	180
	IDATPDIEHD	TKGWSNEKLM	AEVAPDAFSG	NPEGQFFPDS	YEIDAGGSDL	RIYQIAYKAM
orf7-1	130	140	150	160	170	180
	IDATPDIGHD	TKGWSNEKLM	AEVAPDAFSG	NPEGQFFPDS	YEIDAGGSDL	QIYQTAYKAM
orf7a.pep	190	200	210	220	230	240
	QRRLEAWES	RQDGLPYKNP	YEMLIMASLI	EKETGHEADR	DHVASVFVNR	LKIGMRLQTD
orf7-1	190	200	210	220	230	240
	QRRLEAWES	RQDGLPYKNP	YEMLIMASLI	VEKETGHEADR	DHVASVFVNR	LKIGMRLQTD
orf7a.pep	250	260	270	280	290	300
	PSVIYMGAA	YKGKIRKADL	RRDTPYNTYT	RGGLPPTPIA	LPGKAALDAA	AHPSGEKYLY
orf7-1	250	260	270	280	290	300
	PSVIYMGAA	YKGKIRKADL	RRDTPYNTYT	RGGLPPTPIA	LPGKAALDAA	AHPSGEKYLY
orf7a.pep	310	320	330			
	FVSKMDGTGL	SQFSDLTEH	NAAVRKYILK	KX		
orf7-1	310	320	330			
	FVSKMDGTGL	SQFSDLTEH	NAAVRKYILK	KX		

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF7 shows 94.7% identity over a 187aa overlap with a predicted ORF (ORF7.ng) from *N. gonorrhoeae*:

50  
55  
60

orf7	MRGGRPDSVTVQIIEGSRFSHMRKV	IDATPDIGHDTKGWSNEKLM	AEVAPDAFSGNPEGQ	60		
orf7ng	MRGGRPDSVTVQIIEGSRFSHMRKV	IDATPDIGHDTKGWSNEKLM	AEVAPDAFSGNPEGQ	60		
orf7	FFPDSYEIDAGGSDLQIYQTAYKAM	QRRLEAWESRQDGLPYKNPYEMLIMASLI	VEKETG	120		
orf7ng	FFPDSYEIDAGGSDLQIYQTAYKAM	QRRLEAWESRQDGLPYKNPYEMLIMASLI	VEKETG	120		
orf7	HEAXXDHVASVFVNR	LKIGMRLQTD	PSVIYMGAA	YKGKIRKADLRRDTPYNTYT	TRGGGLP	180
orf7ng	HEAXXDHVASVFVNR	LKIGMRLQTD	PSVIYMGAA	YKGKIRKADLRRDTPYNTYT	TRGGGLP	180
orf7	PTPIALP					187

5

1	MRGGRPDSVT	VQIEGSRFS	HMRKVIDATP	DIGHDTKGWS	NEKLMAEVAP
51	DAFSGNPEQG	FFPDSYEIDA	GGSDLQIYQT	AYKAMQRRLN	EAWAGRQDGL
101	PYKNPYEMLI	MASLLEKETG	HEADRDHVAS	VFVNRLKIGM	RLQTDPSVIY
151	GMGAAYKSKI	RKADLRDTP	YNTYTGGLP	PTRIALPGKA	AMDAAAHPSG
201	EKLYLFVSKM	DGTGLRSQFSH	DLTEHNAAVR	KYILKK*	

	1	..taccgaatca	AGATTGCCAA	AAATCAGGGT	ATTTCGTCGG	TCGGCAGGAA
	51	ACTTGCCgaA	GACCGCATCG	TGTTCAGCAG	GCATGTTTTG	ACAGCGGCGG
	101	CCTACGTTTT	GGGTGTGCAC	AACAGGCTGC	ATACGGGGAG	gTACAGATTG
15	151	CCTTCGGAAG	TGTCGTCTTG	GGATATCTTG	CAGAAAATGC	GCGGCGGCAG
	201	CGCGGATTCC	GTTACCGTGC	AGATTATCGA	AGGTTGCGGT	TTCTGCATA
	251	TGAGGAAAGT	CATCGACGCA	ACGCCCGACA	TCGGACACGA	CACCAAAGGC
	301	TGGAGCAATG	AAAAACTGAT	GGCGGAAGTT	GCGCCCGATG	CCTTCAGCGG
	351	CAATCCTGAA	GGGCAGTTTT	TTCCCGACAG	CTACGAAATG	GATGCGGGCG
20	401	GCAGCGATTG	GCAGCTTTAC	CAAAAGCGCT	ACAAGGCGAT	GCAACGCGCG
	451	CTGAACGAGG	CATGGCGAGG	CAGGCAGGAC	GGGCTGCCTT	ATAAAAACCC
	501	TTATGAAATG	CTGATTATGG	CGAGCCTGAT	CGAAAAGGAA	ACGGGGCATG
	551	AGGCCGACCG	CGACCATGTC	GCTTCCGTCT	TCGTCAACCG	CCTGAAAACT
	601	GGTATGCGCG	TGCAAACCGA	CCCGTCCGTG	ATTTACGGCA	TGGGTGCGGC
	651	ATACAAGGGC	AAAATCCGTA	AAGCCGACCT	GCGCCGCGAC	AGCCCTACAA
25	701	aCAcCTAtac	gggcgggggc	ttgcgcgcaa	cccggtattgc	gctgcccggc
	751	Aaggcgggcaa	tggatgccgc	cgccccaccg	tccggcgaaA	aataactgtA
	801	tttcgtgtcC	AAAATGGACG	GCACGGGCTT	GAGCCAGTTC	AGCCATGATT
	851	ATGTCGCAACA	CAACGCGCGc	dTcCGCAAA	TATTTTTGAA	AAAAATAA

```

30      1  ..YRIKIAKNQG ISSVGRKLAE DRIVFSRHVL TAAAYVLGVH NRLHTGTYRL
      51  PSEVSAWDIL QKMRGGRPDS VTVQIIEGSR FSHMRKVIDA TPDIGHDTKG
     101  WSEKLMMAEV APDAFSGNPE GQFFPDSEI DAGGSDLQIY QTAYKAMQRR
     151  LNEAWAGRQD GLPYKNPYEM LIMASLIEKE TGHEADRDRH ASVFENRLKI
     201  GMRQLTDPVS IYMGGAAYKG KIRKADLRD TPYNTYTGGV LPPTRIALPG
35     251  KAAMDAAAHP SGEKLYFVFS KMDGTGLSQF SHDLTEHNAA VRKYILKK*

```

		10	20	30	40	50	60
40	orf7-1.pep	KLLKWSAVFLTVSAAVFAALLFVPKDNGRAYRIKIAK	NOGISSVGRKLAEDRIVFSRHVL				
	orf7ng-1				YRIKIAKNOGISSVGRKLAEDRIVFSRHVL		
					10	20	30
45	orf7-1.pep	70	80	90	100	110	120
	orf7ng-1	TAAAYVLGVHNRLHTGT	YRLPSEVSAWDILOKMRGGRPD	SVTVQII	EGSRFSHMRKVIDA		
		40	50	60	70	80	90
50	orf7-1.pep	130	140	150	160	170	180
	orf7ng-1	TPDIGHDTKGWSNEKLM	AEVAPDAFSGNPEGQFFPD	SYEIDAGGSDLQIYQTAYKAMQRR			
		100	110	120	130	140	150
55	orf7-1.pep	190	200	210	220	230	240
	orf7ng-1	LNEAWESRQDGLPYKNPY	EMLIMASLVEKETGHEADRDH	VASVFVNR	LKIGMRLQTDPSV		
		160	170	180	190	200	210
60	orf7-1.pep	250	260	270	280	290	300
		IYGMGAAYK GKIRKADLR	RDTPYNTYTRGGLPPTPI	ALPGKAALDAAHPSGEKYLYFVS			

sp|P28306|YCEG\_ECOLI HYPOTHETICAL 38.2 KD PROTEIN IN PABC-HOLB INTERGENIC REGION  
gi|1787339 (AE000210) o340; 100% identical to fragment YCEG\_ECOLI SW: P28306 but  
has 97 additional C-terminal residues [Escherichia coli] Length = 340  
Score = 79 (36.2 bits), Expect = 5.0e-57, Sum P(2) = 5.0e-57  
Identities = 20/87 (22%), Positives = 40/87 (45%)

Query: 70 SVTVQIIEGSRFSHRKVIDATPDIGH 96  
++++EG R S K + P I H  
Sbjct: 109 QFPLRLVEGMRLSDYLKQLREAPYIKH 135

Query: 120 EGQFFPDYSYIDAGGSDLQIYQTAYKAMQRRRLNEAWAGRDGLPYKNPYEMLIMASLIEK 179  
EG F+PD++ A +D+ + + A+K M + ++ AW GR DGLPYK+ +++ MAS+IEK  
Sbjct: 158 EGWFWDPTWMTANTTDDVALLKRAHKMKVKAVDSAWEGRADGLPYKDKNQLVMTASLIEK 217

Query: 180 ETGHEADRDHVASVFVNRLLKIGMRLQTDPSVIYMGAAAYKGKIRKADLRDTPYNTYTGG 239  
ET ++RD VASVF+NRL+IGMRLQTD+VIYMG Y GK+ +ADL T YNTYT  
Sbjct: 218 ETAVASERDKVASVFINRLRIGMRLQTDPTVIYMGGERYNGKLSRADLETPTAYNTYIT 277

Query: 240 GLPPTRIALPGKAAMDAAHPSGEKYL FVSKMDG 274  
GLPP IA PG ++ AAHP+ YLYFV+ G  
Sbjct: 278 GLPPGAIATPGADSLKAAHPAKTFYLYFVADGKG 312

### Example 6

	1	CGTTTCAAAA	TGTTAACTGT	GTTGACGGCA	ACCTTGATTG	CCGGACAGGT
	51	ATCTGCCGCC	GGAGGCGGTG	CGGGGGATAT	GAACAGCCG	AAGGAAGTCG
	101	GAAAGGTTTT	CAGAAAGCAG	CAGCGTTACA	GCGAGGAAGA	AATCAAAAAC
50	151	GAACGCGCAC	GGCTTGCGGC	AGTGGGCGAG	CGGGTTAATC	AGATATTTAC
	201	GTTGCTGGGA	GGGGAACCCG	CCTTGCAAAA	GGGGCAGCGC	GGAACGGCTC
	251	TGGCAACCTA	TATGCTGATG	TTGGAACGCA	CAAAATCCCC	CGAAGTCGCC
	301	GAACGCGCCT	TGGAATGGC	CGTGTGCTG	AACCGGTTT	AACAGGCGGA
	351	AATGATTAT	CAGAAATGGC	GGCAGATTGA	CGTATACCG	GGTAAGCGCG
	401	AAAAACGGC	GGGTGGCTG	CGGAACGTGC	TGAGGGAAAG	AGGAAATCAG
55	451	CATCTGGACG	GACGGGAAGA	AGTGCTGGCT	CAGGCGGACG	AAGGACAG

```

1  ..RFKMLTVLTA TLIAGQVSAA GGGAGDMKQP KEVGKVFRRKQ QRYSEEEIKN
51 ERARLAAVGE RVNQIFTLG GGTALQKGQA GTALATYMLM LERTKSPEVA
101 ERALEMAVSL NAFEQAEMII QKWRQIEPIP GKAQKRAEWL RNVLRERGNQ

```

1	ATGTTACCTA	ACCGTTTCAA	AATGTTAACT	GTGTTGACGG	CAACCTTGAT
51	TGCCGGACAG	GTATCTGCCG	CCGGAGGCGG	TGCGGGGGAT	ATGAAACAGC
101	CGAAGGAAGT	CGGAAAGGTT	TTCAGAAAGC	AGCAGCGTTA	CAGCGAGGAA
151	GAAATCAAAA	ACGAAACGCG	ACGGCTTGCG	CGAGTTGGCG	ACGGGGTTAA
201	TCAGATATTT	ACGTTGCTGG	GAGGGGAAAC	CGCCTTGCAA	AAGGGGCAAG
251	CGGGAACGGC	TCTGGCAACC	TATATGCTGA	TGTTGGAACG	CACAAAATCC
301	CCCGAAGTCG	CCGAACGCGC	CTTGGAATG	GCCGTGTGCG	TGAACGCGTT
351	TGAACAGGCG	GAAATGATTT	ATCAGAAATG	CGCGCAGATT	GAGCCTATAC
401	CGGGTAAAGC	CGAAACACGG	CGGGGTGGC	TGCGGAACGT	GCTGAGGGAA
451	AGAGGAAATC	AGCATCTGGA	CGGACTGGAA	GAAGTGCTGG	CTCAGGCGGA
501	CGAAGGACAG	AACCGCAGGG	TGTTTTTATT	GTTGGCACAA	GCCGCCGTGC
551	AACAGGACGG	GTTGGCGCAA	AAAGCATCGA	AAGCGGTTG	CCGCGCGCGC
601	TTGAAATATG	AACATCTGCC	CGAAGCGGCG	GTTGCCGATG	TGGTGTTCAG
651	CGTACAGGGA	CGCGAAAGAG	AAAAGGCAAT	CGGAGCTTTG	CAGCGTTTTG
701	CGAAGCTCGA	TACGGAAATA	TTGCCCCCCA	CTTTAATGAC	GTTGCGTCTG
751	ACTGCACGCA	AATATCCCGA	AATACTCGAC	GGCTTTTTTC	AGCAGACAGA
801	CACCCAAAAC	CTTTCGGCCG	TCTGGCAGGA	AATGGAATTT	ATGAATCTGG
851	TTTCCCTGCA	CAGGCTGGAT	GATGCTTATG	CGCGTTTGAA	CGTGCTGTGT
901	GAACGCAATC	CGAATCGAGA	CCTGTATATT	CAGGCAGCGA	TATTGGCGGC
951	AAACCGAAAA	GAAGGTGCTT	CCGTTATCGA	CGGCTACGCC	GAAAAGGCAT
1001	ACGGCAGGGG	GACGGAGGAA	CAGCGGAGCA	GGGCGGCGCT	AACGGCGGCG
1051	ATGATGTATG	CCGACCCGAG	GGATTACGCC	AAAGTCAGGC	AGTGGCTGAA
1101	AAAAGTATCC	CGCCCGGAAT	ACCTGTTCGA	CAAAGGTGTG	CTGGCGGCTG
1151	CGGCGGCTGT	CGAGTTGGAC	GGCGGCAGGG	CGGCTTTGCG	GCAGATCGGC
1201	AGGGTGCGGA	AACTTCCCGA	ACAGCAGGGG	CGGTATTTTA	CGGCAGACAA
1251	TTTGTCCAAA	ATACAGATGC	TCGCCCTGTC	GAAGTCCGCC	GATAAACCGG
1301	AGGCTTTTGA	GGGGTTGGAC	AAGATTATCG	AAAAACCGCG	TGCCCGCAGT
1351	AATACAGAGT	TACAGGCAGA	GGCATTGGTA	CAGCGGTCAG	TTGTTTACGA
1401	TCGGCTTGGC	AAGCGGAAAA	AAATGATTTT	AGATCTTGAA	AGGGCGTTCA
1451	GGCTTGCACC	CGATAACGCT	CAGATTATGA	ATAATCTGGG	CTACAGCCTG
1501	CTGACCGATT	CCAAACGTTT	GGACGAAGTT	TTCCGCCCTG	TTACAGCGGC
1551	ATACCAAATC	ACACCGGACG	ATACCGCTGT	CACACGACAG	ATAGGCTGGG
1601	CGTATTACCT	GAAAGGCGAC	GCGGAAAGCG	CGCTGCCGTA	TCTGCGGTAT
1651	TGTTTTGAAA	ACGACCCCGA	GCCCGAAGTT	GCCGCCCAT	TGGGCGAAGT
1701	GTTGTGGGCA	TTGGGCGAAC	GCGATCAGGC	GGTTGACGTA	TGGACCGCAG
1751	CGGCACACCT	TACGGGAGAC	AAGAAAATAT	GGCGGGAAAC	GCTCAAACGT
1801	CACGGCATCG	CATTGGCCCA	ACCTTCCCGA	AAACCTCGGA	AATAA

	1	MLPNRFKMLT	VLATATLIAGQ	VSAAGGGAGD	MKPQKEVGKV	FRKQORYSEE
	51	EIKNERARLA	AVGERVNQIF	TLLGGETALQ	KGQAGTALAT	YMLMLERTKS
	101	PEVAERALEM	AVSLNAFEQA	EMIYQKWRQI	EPIPGKAQKR	AGWLRNVLR
45	151	RGNQHLDGLE	EVLAQADEGQ	NRRVFLLLAQ	AARQDGLAQ	KASKAVRRAA
	201	LKYEHLPEAA	VADVVSFVQG	REKEKAIGAL	QLRALDTEI	LPPTLMTLRL
	251	TARKYPEILD	GFFEQTDTQN	LSAVWQEMEI	MNLVSLHRLD	DAYARLNVLL
	301	ERNPNADLYI	QAAILAANRK	EGASVIDGYA	EKAYGRGTEE	QRSRAALTAA
	351	MMYADRRDYA	KVRQWLKKVS	APEYLFDKGV	LAAAAAVELD	GGRAALRQIG
	401	RVRKLPQQQG	RYFTADNLSK	IQMLALSKLP	DKREALRGLD	KIIEKPPAGS
50	451	NTELQAEALV	QRSVVYDRLG	KRKKMISDLE	RAFRLAPDNA	QIMNNLGYSL
	501	LTDSKRLDEG	FALLQTAYQI	NPDDTAVNDS	IGWAYYLKGD	AESALPYLRY
	551	SFENDPEPEV	A AHLGEVLWA	LGERDQAVDV	WTQAAHLTGD	KKIWRETLKR
	601	HGIALPOPSR	KPRK*			

55 Homology with a predicted ORF from *N.meningitidis* (strain A)

**60**

```

orf9.pcp      RFKMLTVLTATILLAGQVSAAGGGAGDMDRQPKEVGKVFRKOORYSEEEIKNERARLA
               || :||::|:||:|||| |:||:| ||||| ||||| ||||| |||||
orf9a         MLPARETII SVLAAALLAGOAYAA--GAADAKPPKEVGKVFRKOORYSEEEIKNERARLA

```

-80-

		10	20	30	40	50
		60	70	80	90	100
5	orf9.pep	AVGERVNQIFTLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA				
	orf9a	AVGERVNQIFTLGXETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA				
		60	70	80	90	100
10	orf9.pep	120	130	140	150	160
	orf9a	EMIQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGREVLQAQADEGQ				
		120	130	140	150	160
15	orf9a	AAVQDGLAQKASKAVRRAALRYEHLPEAAVADVFSVQXREKEKAIGALQRLAKLDTETI				
		180	190	200	210	230

The complete length ORF9a nucleotide sequence <SEQ ID 43> is:

	1	ATGTTACCCG	CCCGTTTCAC	CATTTTATCT	GTGCTCGCGG	CAGCCCTGCT
	51	TGCCGGGCAG	GCGTATGCCG	CCGGCGCGGC	GGATGCGAAG	CCGCCGAAGG
20	101	AAGTCGGAAG	GGTTTTTACA	AAGCAGCAGC	GTTACAGCGA	GGAAGAAATC
	151	AAAAACGAAC	GCGCACGGCT	TGCGGCAGTG	GGCGAGCGGG	TTAATCAGAT
	201	ATTTACGTTG	CTGGGANGGG	AAACCGCCTT	GCAAAAAGGGG	CAGGCGGGAA
	251	CGGCTCTGGC	AACCTATATG	CTGATGTTGG	AACGCACAAA	ATCCCCCGAA
	301	GTCGCCGAAC	GCGCCTTGGA	AATGGCCGTG	TCNCTGAACG	CGTTTGAACA
25	351	GCGCGAAATG	ATTTATCAGA	AATGGCGGCA	GATTGAGCCT	ATACCGGGTA
	401	AGGCGCAAAA	ACGGGCGGGG	TGGCTGCGGA	ACGTGCTGAG	GGAAAGAGGA
	451	AATCAGCATC	TAGACGGACT	GGAAGAANTG	CTGGCTCAGG	CGGACGAANG
	501	ACAGAACCGC	AGGGTGT TTT	TATTGTTGGC	ACAAGCCGCC	GTGCAACAGG
	551	ACGGGTTGGC	GCAAAAAGCA	TCGAAAGCGG	TTCGCCGCGC	GGCGTTGAGA
30	601	TATGAACATC	TGCCCGAAGC	GGCGGTTGCC	GATGTGGTGT	TCAGCGTACA
	651	GGNACGCGAA	AAGGAAAAGG	CAATCGGAGC	TTTGCAGCGT	TTGGCGAAGC
	701	TCGATACGGA	AATATTGCCC	CCCACTTTAA	TGACGTTGCG	TCTGACTGCA
	751	CGCAAATATC	CCGAAATACT	CGACGGCTTT	TTTCAGCAGA	CAGACACCCA
	801	AAACCTTTTC	GCCGCTGGC	AGGAAATGGA	AATTATGAAT	CTGGTTTCCC
35	851	TGCACAGGCT	GGATGATGCC	TATGCGCGTT	TGAACGTGCT	GTTGGAACGC
	901	AATCCGAATG	CAGACCTGTA	TATTCAGGCA	GCGATATTGG	CGGCAACCGG
	951	AAAAGAANGT	GCTTCCGTTA	TCGACGGCTA	CGCCGAAAAG	GCATACGGCA
	1001	GGGGGACGGG	GGAACAGCGG	GGCAGGGCGG	CAATGACGGC	GGCGATGATA
	1051	TATGCCGACC	GAAGGGATTA	CACCAAAGTC	AGGCAGTGGT	TGAAAAAAGT
40	1101	GTCCGCGCCG	GAATACCTGT	TCGACAAAGG	TGTGCTGGCG	GCTGCGCGCG
	1151	CTGTGCGAGT	GGACNGCGGC	AGGGCGGCTT	TGCGGCAGAT	CGGCAGGGTG
	1201	CGGAAACTTC	CCGAACAGCA	GGGGCGGTAT	TTTACGGCAG	ACAATTTGTC
	1251	CAAAATACAG	ATGTTTCGCC	TGTCGAAGCT	GCCCCACAAA	CGGGAGGCTT
	1301	TGAGGGGGTT	GGACAAGATT	ATCGAAAAAC	CGCCTGCCGG	CAGTAATACA
45	1351	GAGTTACAGG	CAGAGGCATT	GGTACAGCGG	TCAGTTGTTT	ACGATCGGCT
	1401	TGGCAAGCGG	AAAAAATGA	TTTCAGATCT	TGAAAGGGCG	TTTACGGCTG
	1451	CACCCGATAA	CGCTCAGATT	ATGAATAATC	TGGGTACAG	CCTGCTTTCC
	1501	GATTCCAAAC	GTTTGGACGA	AGGCTTCGCC	CTGCTTCAGA	CGGCATACCA
	1551	AATCAACCCG	GACGATACCG	CTGTCAACGA	CAGCATAGGC	TGGGCGTATT
50	1601	ACCTGAAANG	CGACGCGGAA	AGCGCGCTGC	CGTATCTGCG	GTATTCTGTT
	1651	GAAAACGACC	CCGAGCCCGA	AGTTGCCGCC	CATTGCGGCG	AAGTGTGTGT
	1701	GGCATTGGGC	GAACGCGATC	AGGCGGTTGA	CGTATGGACG	CAGGCGGCAC
	1751	ACCTTACGGG	AGACAAGAAA	ATATGGCGGG	AAACGCTCAA	ACGTCACGGC
	1801	ATCGCATTGC	CCCAACCTTC	CCGAAAACCT	CGGAAATAA	

55 This encodes a protein having amino acid sequence <SEQ ID 44>:

	1	MLPARFTILS	VLAAALLAGO	AYAAGAADAK	PPKEVGKVER	KQORYSEEEI
	51	KNERARLAAV	GERVNQIFTL	LGXETALQKG	QAGTALATYM	LMLERTKSPE
	101	VAERALEMAV	SLNAFEQAEM	IYQKWRQIEP	IPGKAQKRAG	WLRNVLRERG
	151	NQHLGLEEX	LAQADEXQNR	RVFLLLAQAA	VOODGLAQA	SKAVRRAALR
60	201	YEHLPAAVA	DVVFVQXRE	KEKAIGALQR	LAKLDTEILP	PTLMTLRLTA
	251	RKYPEILDGF	FEQTDQNL	AVWQEMEIMN	LVSLHRLDDA	YARLNVLLER
	301	NPNADLYIQ	AILAANRKEK	ASVIDGYAEK	AYGRGTGEQR	GRAAMTAAMI
	351	YADRRDYTKV	RQWLKVSAP	EYLFDKGVLA	AAAAVELDXG	RAALRQIGRV
	401	RKLPEQQGRY	FTADNLSKI	MFALSKLPDK	REALRGLDKI	IEKPPAGSNT
65	451	ELQAEALVQR	SVVYDRLGKR	KKMISDLERA	FRLAPDNAQI	MNNLGYSLLS
	501	DSKRLDEGFA	LLQTAYQINP	DDTAVNDSIG	WAYYLKXDAE	SALPYLRYSF
	551	ENDPEPEVAA	HLGEVLWALG	ERDQAVDVWT	QAAHLTGDKK	IWRETLKRHG



601 IALPQPSRKPK\*

ORF9a and ORF9-1 show 95.3% identity in 614 aa overlap:

5	orf9a.pep	10 20 30 40 50	MLPARFTILSVLAAALLAGQAYAAG--AADAKPPKEVGKVFRRKQORYSEEEIKNERARLA
	orf9-1	10 20 30 40 50 60	MLPNRFKMLTVLTATLIAGQVSAAGGGAGDMKQPKVEVGKVFRRKQORYSEEEIKNERARLA
10	orf9a.pep	60 70 80 90 100 110	AVGERVNIQIFTLGKETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
	orf9-1	60 70 80 90 100 110 120	AVGERVNIQIFTLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
15	orf9a.pep	120 130 140 150 160 170	EMIQKWRQIEPIPGKAQKQAGWLRNVLRERGNQHLDGLEEXLAQADEXQNRVFLLLAQ
	orf9-1	120 130 140 150 160 170 180	EMIQKWRQIEPIPGKAQKQAGWLRNVLRERGNQHLDGLEEVLAQADEGQNRVFLLLAQ
20	orf9a.pep	180 190 200 210 220 230	AAVQQDGLAQKASKAVRRAALRYEHLPEAAVADVVSQVQREKEKAIGALQRLAKLDTEI
	orf9-1	180 190 200 210 220 230 240	AAVQQDGLAQKASKAVRRAALKYEHLPEAAVADVVSQVQREKEKAIGALQRLAKLDTEI
25	orf9a.pep	240 250 260 270 280 290	LPPTLMTLRLTARKYPEILDGFFEQTDTQNL SAVWQEMEIMNLVSLHRLDDAYARLNVLL
	orf9-1	240 250 260 270 280 290 300	LPPTLMTLRLTARKYPEILDGFFEQTDTQNL SAVWQEMEIMNLVSLHRLDDAYARLNVLL
30	orf9a.pep	300 310 320 330 340 350	ERNPNADLYIQAAILAANRKEKXASVIDGYAEKAYGRGTGEQGRGAAMTAAMIYADRRDYT
	orf9-1	300 310 320 330 340 350 360	ERNPNADLYIQAAILAANRKEGASVIDGYAEKAYGRGTGEQGRGAAMTAAMMYADRRDYA
35	orf9a.pep	360 370 380 390 400 410	KVRQWLKKSVAPEYLFDKGVLA AAAA VELDGGRAALRQIGRVRKLPEQQGRYFTADNLSK
	orf9-1	360 370 380 390 400 410 420	KVRQWLKKSVAPEYLFDKGVLA AAAA VELDGGRAALRQIGRVRKLPEQQGRYFTADNLSK
40	orf9a.pep	420 430 440 450 460 470	IQMFALSKLPDKREALRGLDKIIEKPPAGSNTLQAEALVQSVVYDRLGKRKKMISDLE
	orf9-1	420 430 440 450 460 470 480	IQMLALSKLPDKREALRGLDKIIEKPPAGSNTLQAEALVQSVVYDRLGKRKKMISDLE
45	orf9a.pep	480 490 500 510 520 530	RAFRLAPDNAQIMNNLGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKXD
	orf9-1	480 490 500 510 520 530 540	RAFRLAPDNAQIMNNLGYSLLTDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGD
50	orf9a.pep	540 550 560 570 580 590	AESALPYLRYSFENDPEPEVAHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWRETLKR
	orf9-1	540 550 560 570 580 590 600	AESALPYLRYSFENDPEPEVAHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWRETLKR
55	orf9a.pep	600 610	HGIALPQPSRKPKKX
	orf9-1	600 610	HGIALPQPSRKPKKX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF9 shows 82.8% identity over a 163aa overlap with a predicted ORF (ORF9.ng) from *N.*

*gonorrhoeae:*

	Orf9	RFKMLTVLTATLIAGQVSAAGGGAGDMKPKEVGKVFRKKORYSEEEIKNERAR	54
5	orf9ng	MIMLPARFTILSVLAAALLAGQAYAA--GAADVELPKEVGKVLRKHRRYSEEEIKNERAR	58
	orf9	LAAVERVNQIFTLGGETALQKGAGTALATYMLMLERTKSPEVAERALEMAVSLNAFE	114
10	orf9ng	LAAVERVNRVFTLLGGETALQKGAGTALATYMLMLERTKSPEVAERALEMAVSLNAFE	118
	orf9	QAEMIQKWQRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGREEVLAQADEGQ	166
	orf9ng	QAEMIQKWQRQIEPIPGAQPKAGWLRNVLKEGGNPHLDRLEEVPAQSDYVHQPMIFLLL	178

15 The ORF9ng nucleotide sequence <SEQ ID 45> was predicted to encode a protein having including acid sequence <SEQ ID 46>:

20

1	MIMLPARETI	LSVLAAALLA	GOAYAAGAAD	VELPKEVGKV	LRKHRRYSEE
51	EIKNERARLA	AVGERVNRVF	TLLGGETALQ	KGQAGTALAT	YMLMLERTKS
101	PEVAERALEM	AVSLNAFEQA	EMIYQKWRQI	EPIPGEAQKP	AGWLRNLVLE
151	GGNPHLDRLR	EVPAQSDYVH	QPMIFLLLVQ	AAVQHGVAQ	KPSKAVRPAA
201	YNYEVLPETA	GADAVFCVQG	PQYEKATQSF	PCFGRNPQTE	NIAPPFNELF
251	RPTARPISPK	LLQRFFRTEP	NLAKPFRPPG	PEMETYQTGF	PRPLTRNNPT

Amino acids 1-28 are a putative leader sequence, and 173-189 are predicted to be a transmembrane domain.

25 Further sequence analysis revealed the complete length ORF9ng DNA sequence <SEQ ID 47>>:

	1	ATGTTACCCG	CCCGTTTCAC	TATTTTATCT	GTCCCTCGCAG	CAGCCCTGCT
	51	TGCCGGACAG	GCGTATGCTG	CCGGCGCGGC	GGATGTGGAG	CTGCCGAAGG
	101	AAGTCGGAAG	GGTTTAAAGG	AAACATCGGC	GTTACAGCGA	GGAAGAAATC
	151	AAAAACGAAC	CGCGACGGCT	TGCGGCAGTG	GGCGAACGGG	TCAACAGGGT
30	201	GTTTACGTTG	TTGGCGGGTG	AAACGGCTTT	GCAGAAAGGG	CAGGCGGGAA
	251	CGGCTCTGGC	AACCTATATG	CTGATGTTGG	AACGCACAAA	ATCCCCCGAA
	301	GTCGCGGAAC	GCGCCTTGGA	AATGGCCGTG	TCGCTGAACG	CGTTTGAACA
	351	GGCGGAAATG	ATTTATCAGA	AATGcgcgca	gacgagcct	ataCggggtg
	401	aggcgcaaaa	accgGcgggG	tggctgcgga	acgtattgaa	ggaagggGga
35	451	aaTCAGCATC	TGGAcggggt	gaaagaggTG	CtggcgcaAT	cggacgatGT
	501	GCAAAAAcgc	aggaTATTTT	TGCTGCTGGT	GCAAGCCGCG	GTGcagcagg
	551	gTGGGGTGGC	TCAAAAAGCA	TCGAAAGCGG	TTGCGcgtgc	GGcgttgaAG
	601	TATGAACATC	TGCCgaagc	ggcggTTGCC	GATGcggTGT	TCGGCGTACA
	651	GGAGCGCGAA	AAGGAAAagg	caaTCGAAGC	TTTGCAGCGT	TTGGCGAAGC
40	701	TCGATACGGA	AATATTGCCC	CCCACTTTAA	TGACGTTGCG	TCGACTGCA
	751	CGCAAAATATC	CCGAAATACT	CGACGGCTTT	TTGCGAGCAG	CAGACACCCA
	801	AAACCTTTCG	CGCGTCTGGC	AGGAAATGGA	AATTATGAAT	CTGGTTTCCC
	851	TGCGTAAGCC	GGATGATGCC	TATGCGCGTT	TGAACGTGGT	GTTGGAACAC
	901	AACCCGAATG	CAAACCTGTA	TATTcAGCGG	CGCATATTGG	CGGCAAACCG
45	951	AAAAGAAGGT	GCGTCCGTTA	TCGACGGCTA	CGCCGAAAGG	GCATACGGGA
	1001	GGGGGACGGG	GCACAGCGGG	GGCagggcgg	caATgacggc	GGCGATGATA
	1051	TATGCCGACC	GCAGGGATTA	CGCCAAAGTC	AGGCAGTGGT	TGAAAAAAGT
	1101	GTCCGCGCCG	GAATACCTGT	TCGACAAAGG	CGTGCTGGCG	GCTGCGCGGG
	1151	CTGCCGAATT	GGACGGAGCG	CGGGCGGCTT	TGCGGCAGAT	CGGCGAGGGT
50	1201	CGGAAACTTC	CCGAACAGCA	GGGGCGGTAT	TTTACGGCAG	ACAAATTGTC
	1251	CAAAATACAG	ATGCTCGCCC	TGTCGAAGCT	GCCCCACAAA	CGGGAAGCCC
	1301	TGATCGGGCT	GAACAACATC	ATCGCCAAAC	TTTCCGGCGC	GGGAAGCACG
	1351	GAACCTTTGG	CGGAAGCATT	GGCACAAGCT	TTCATTATTT	ACGaacAGTT
	1401	cggCAAAACGG	GGA AAAATGA	TTGGCGACCT	tgaAACcgcg	CTCAAACCTTA
	1451	CGCCCCGATA	TGCACAAATT	ATGAATAATC	TGGGCTACAG	CCTGCTTTCC
	1501	GATTCCAAAC	GTTTTGGACGA	GGGTTTCGCC	CTGCTTCAGA	CGGCATACCA
	1551	AATCAACCCG	GACGATACCG	CCGTTTAACGA	CAGCATAGGC	TGGGCTATTT
	1601	ACCTGAAAGG	CGACGcggaA	AGCGCGCTGC	CGTATCTGcg	gtattcgttt
55	1651	gAAAACGACC	CCGAGCCCGA	AGTTGCCGCC	CATTTGGGCG	AAGTGTGTGT

1701 GGCATTGGGC GAACGCGATC AGGCGGTTGA CGTATGGACG CAGGCGGCAC  
 1751 ACCTTAGGGG AGACAAGAAA ATATGGCGGG AGACGCTCAA ACGCTACGGA  
 1801 ATCGCCTTGC CCGAGCCTTC CCGAAAACCC CGGAAATAA

This encodes a protein having amino acid sequence <SEQ ID 48>:

5           1   MLPARFTILS VLAAALLAQ AYAAGAADVE LPKEVGKVLK KHRRYSEEEI  
           51   KNERARLA AV GERVNRVFTL LGGETALQKG QAGTALATYM LMLERTKSPE  
 101       VAERALEMAV SLNAFEQAEM IYQKWRQIEP IPGEAQKPAG WLRNVLKEGG  
 151       NQHLDELKEV LAQSDDVQKR RIFLLLVQAA VQGGGVAQKA SKAVRRAALK  
 201       YEHLPEAAVA DAVFGVQGRE KEKAIEALQR LAKLDTEILP PTLMTLRLTA  
 251       RKYPEILDGF FEQTDTONLS AVWQEMEIMN LVSLRKPDDA YARLNVLEH  
 301       NPNANLYIQA AILAAANRKEG ASVIDGYAEK AYGRGTGEQR GRAAMTAAMI  
 351       YADRRDYAKV RQWLKKVSAP EYLFDKGVLA AAAAAELDGG RAALRQIGRV  
 401       RKLPEQQGRY FTADNLSKIQ MLALSCLPKD REALIGLNNI IAKLSAAGST  
 451       EPLAEALAQR SIIYEQFGKR GKMIADLETA LKLPDQNAI MNNLGYSLLS  
 501       DSKRLDEGFA LLQTAYQINP DDTAVNDSIG WAYYLLKGD AE SALPYLRYSF  
 551       ENDPEPEVAA HLGEVLWALG ERDQAVDVWT QAAHLRGDKK IWRETLKRYG  
 601       IALPEPSRKP RK\*

ORF9ng and ORF9-1 show 88.1% identity in 614 aa overlap:

20	orf9-1.pep	MLPNRFKMLTVLTATLIAGQVSAAGGAGDMKQPKVEVGKVERKQORYSEEEIKNERARLA	10	20	30	40	50	60
	orf9ng-1	MLPARFTILSVLAAALLAQAYAG--AADVELPKEVGKVLKHHRRYSEEEIKNERARLA	10	20	30	40	50	
25	orf9-1.pep	AVGERVNIQIFTLGGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA	70	80	90	100	110	120
	orf9ng-1	AVGERVNRVFTLLGGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA	60	70	80	90	100	110
30	orf9-1.pep	EMIIYQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGLEEVLAQADEGQNRVFLLLAQ	130	140	150	160	170	180
	orf9ng-1	EMIIYQKWRQIEPIPGEAQKPAGWLRNVLKEGGNQHLDGGLKEVLAQSDDVQKRRIFLLLVQ	120	130	140	150	160	170
35	orf9-1.pep	AAVQQDGLAQKASKAVRRAALKYEHLPAAVADVFSVQGREKEKAIGALQRLAKLDTEI	190	200	210	220	230	240
	orf9ng-1	AAVQQGGVAQKASKAVRRAALKYEHLPAAVADAVFGVQGREKEKAIEALQRLAKLDTEI	180	190	200	210	220	230
40	orf9-1.pep	LPPTLMTLRLTARKYPEILDGFFEQTDTONLSAVWQEMEIMNLVSLHRLDDAYARLNVLL	250	260	270	280	290	300
	orf9ng-1	LPPTLMTLRLTARKYPEILDGFFEQTDTONLSAVWQEMEIMNLVSLRKPDDAYARLNVLL	240	250	260	270	280	290
45	orf9-1.pep	ERNPNADLYIQAAAILAANRKEGASVIDGYAEKAYGRGTGEQRSAALTAAMMYADRRDYA	310	320	330	340	350	360
	orf9ng-1	EHNPANLYIQAAAILAANRKEGASVIDGYAEKAYGRGTGEQRGRAAMTAAMIYADRRDYA	300	310	320	330	340	350
50	orf9-1.pep	KVRQWLKKVSAPEYLFDKGVLA AAAA AVELDGGRAALRQIGRVKRLPEQQGRYFTADNLSK	370	380	390	400	410	420
	orf9ng-1	KVRQWLKKVSAPEYLFDKGVLA AAAA AVELDGGRAALRQIGRVKRLPEQQGRYFTADNLSK	360	370	380	390	400	410
55	orf9-1.pep	IQMLALSCLPKREALRGLDKIIEKPPAGSNTLQAEALVQSVVYDRLGKRRKMISDLE	430	440	450	460	470	480
	orf9ng-1	IQMLALSCLPKREALIGLNNIIAKLSAAGSTEPLAEALAQR SIIYEQFGKRGKMIADLE	420	430	440	450	460	470
60			490	500	510	520	530	540

```

550      560      570      580      590      600
orf9-1.pep AESALPYLRYSFENDPEPEVA AHLGEVLWALGERDQAVDVVTQA AHLTGDKKIWRETLKR
|||
orf9ng-1    AESALPYLRYSFENDPEPEVA AHLGEVLWALGERDQAVDVVTQA AHLRGDKKIWRETLKR
540      550      560      570      580      590

```

10

15

20

25

30

35

40

45

50

55

60

65

70

Sbjct: 335 GNYEDAKRLIEKAKVLA----PDKKEILFLEADYYSKTKQYDKALEILKKLEKDYPNDSR 390  
 Query: 460 ----RSIIYEQFGKRGKMIADLETALKLTPDNAQIMNNLGYSLLS--DSKRLDEGFALLQ 513  
           +I+Y+ G L A++L P+N N LGYSLL +R++E L++  
 5 Sbjct: 391 VYFMEAIVYDNLGDIKNAEKALRKAIELDPENPDYNYLGYSLLLWYGKERVEEAEELIK 450  
 Query: 514 TAYQINPDDTAVNDSIGWAYYLKGAESALPYLRYSF-ENDPEPEVA AHLGEVLWALGER 572  
           A + +P++ A DS+GW YYLKG D E A+ YL + E +P V H+G+VL +G +  
 10 Sbjct: 451 KALEKDPENPAYIDSMGWVYYLKG DYERAMQYLLKALREAYDDPVVNEHVGDVLLKMGYK 510  
 Query: 573 DQAVDVWTQAAHLRGDKK 590  
           ++A + + +A L + K  
 Sbjct: 511 EEARNYYERALKLLEEGK 528

- 15 Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 7

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 49>:

20 1 AACCTCTACG CCGGCCCGCA GACCACATCC GTCATCGCAA ACATCGCCGA  
 51 CAACCTGCAA CTGGCCAAAG ACTACGGCAA AGTACACTGG TTCGCCTCCC  
 101 CGCTCTTCTG GCTCCTGAAC CAACTGCACA ACATCATCGG CAACTGGGGC  
 151 TGGGCGATTA TCGTTTTAAC CATCATCGTC AAAGCCGTAC TGTATCCATT  
 201 GACCAACGCC TCTTACCGCT CTATGGCGAA AATGCGTGCC GCCGCACCCA  
 25 251 AACTGCAAGC CATCAAAGAG AAATACGGCG ACGACCGTAT GCGCGAACAA  
 301 CAGGCGATGA TGCAGCTTTA CACAGACGAG AAAATCAACC CG<sub>2</sub>CTGGGCG  
 351 GCTGCTGCC TATGCTGTG CAAATCCCCG TCTTCATCGG ATTGTATTGG  
 401 GCATTGTTTC CCTCCGTAGA ATTGCGCCAG GCACCTTGGC TGGGTTGGAT  
 451 TACCGACCTC AGCCGCGCCG ACCCCTACTA CATCCTGCCC ATCATTATGG  
 501 CGGCAACGAT GTTCGCCCAA ACTTATCTGA ACCCGCCGCC GACCGACCCG  
 30 551 ATGCagGCGA AAATGATGAA AATCATGCCG TTGGTTTTCT CsGwCr-TGTT  
 601 CTTCTTCTTC CCTGCCGGks TGGTATTGTA CTGGGTAGTC AACAACTCC  
 651 TGACCATCGC CCAGCAATGG CACATCAACC GCAGCATCGA AAAACAACGC  
 701 GCCCAAGGCG AAGTCGTTTC CTAA

This corresponds to the amino acid sequence <SEQ ID 50; ORF11>:

35 1 ..NLYAGPQTTS VIANIADNLQ LAKDYGKVHW FASPLEWLLN QLHNIIGNWG  
 51 WAIIVLTIIV KAVLYPLTNA SYRSMAMRA AAPKLQAIKE KYGDDRMAQQ  
 101 QAMMQLYTDE KINPLGGCLP MLLQIPVFIG LYWALFASVE LRQAPWLGWI  
 151 TDLSRADPHY ILPIIMAATM FAQTYLNPPP TDPMQAKMMK IMPLVFSXXF  
 201 FFFPAGXVLY WVVNNLLTIA QQWHINRSIE KQRAQGEVVS \*

- 40 Further sequence analysis revealed the complete DNA sequence <SEQ ID 51>:

1 ATGGATTTTA AAAGACTCAC GCGGTTTTTC GCCATCGCGC TGGTGATTAT  
 51 GATCGGCTGG GAAAAGATGT TCCCCACTCC GAAGCCAGTC CCCGCGCCCC  
 101 AACAGGCAGC ACAACAACAG GCCGTAACCG CTCCGCGCGA AGCCGCGCTC  
 151 GCGCCCGCAA CGCCGATTAC CGTAACGACC GACACGGTTC AAGCCGTCAT  
 45 201 TGATGAAAAA AGCGGCGACC TGCGCCGGCT GACCCTGCTC AAATACAAAG  
 251 CAACCGGCGA CGAAAATAAA CCGTTCATCC TGTTTGCGCA CGGCAAGAA  
 301 TACACCTACG TCGCCCAATC CGAACTTTTC GACGCGCAGG GCAACAACAT  
 351 TCTAAAAGGC ATCGGCTTTA GCGCACCGAA AAAACAGTAC AGCTTGGAAG  
 401 GCGACAAAGT TGAAGTCCGC CTGAGCGCGC CTGAAACACG CCGTCTGAAA  
 50 451 ATCGACAAAG TTTATACTTT CACCAAGGC AGCTATCTGG TCAACGTCCG  
 501 CTTGACATC GCCAACGGCA GCGGTCAAAC CGCCAACCTG AGCGCGGACT  
 551 ACCGCATCGT CCGCGACCAC AGCGAACCCG AGGGTCAAGG TTAATTAC  
 601 CACTCTTACG TCGGCCCTGT TGTATTATACC CCTGAAGGCA ACTTCCAAAA  
 651 AGTCAGCTTT TCCGACTTGG ACGACGATGC CAAATCCGGC AAATCCGAGG  
 55 701 CCGAATACAT CCGCAAAACC CCGACCGGCT GGCTCGGCAT GATTGAACAC  
 751 CACTTCATGT CCACCTGGAT TCTCCAACCT AAAGGCAGAG AAAGCGTTTG  
 801 CGCCGCGAGC GAGTGCAACA TCGACATCAA ACGCCGCAAC GACAAGCTGT  
 851 ACAGCACCAG CGTCAGCGTG CCTTTAGCCG CCATCCAAAA CGGCGCGAAA  
 901 GCCGAAGCCT CCATCAACCT CTACGCGCGC CCGCAGACCA CATCCGTCAT  
 60 951 CGCAACATC GCGGACAACC TGCAACTGGC CAAAGACTAC GGCAAGTAC

1001 ACTGGTTCGC CTCCTCGCTC TTCTGGCTCC TGAACCAACT GCACAACATC  
 1051 ATCGGCAACT GGGGCTGGGC GATTATCGTT TTAACCATCA TCGTCAAAGC  
 1101 CGTACTGTAT CCATTGACCA ACGCCTCTTA CCGCTCTATG GCGAAAATGC  
 1151 GTGCCGCCGC ACCCAAATG CAAGCCATCA AAGAGAAATA CGGCGACGAC  
 1201 CGTATGGCGC AACAACAGGC GATGATGCAG CTTTACACAG ACGAGAAAAT  
 1251 CAACCCGCTG GCGGCTGCC TGCCTATGCT GTTGCAAATC CCCGTCTTCA  
 1301 TCGGATTGTA TTGGGCATTG TTCGCCTCCG TAGAATTGCG CCAGGCACCT  
 1351 TGGCTGGGTT GGATTACCGA CCTCAGCCGC GCCGACCCCT ACTACATCCT  
 1401 GCCCATCATT ATGGCGGCAA CGATGTTCGC CCAAACCTTAT CTGAACCCGC  
 1451 CGCCGACCGA CCCGATGCAG GCGAAAATGA TGAATATCAT GCCGTGGTT  
 1501 TTCTCCGTCA TGTTCTTCTT CTTCCCTGCC GGTCTGGTAT TGTACTGGGT  
 1551 AGTCAACAAC CTCCTGACCA TCGCCAGCA ATGGCACATC AACCGCAGCA  
 1601 TCGAAAAACA ACGCGCCAA GCGAAGTCG TTTCCTAA

This corresponds to the amino acid sequence <SEQ ID 52; ORF11-1>:

15 1 MDEFKRLTAFF AIALVIMIGW EKMFPTEPKV PAPOQAAQQQ AVTASAEAL  
 51 APATPITVTT DTVQAVIDEK SGDLRRLTLL KYKATGDENK PFILFGDGKE  
 101 YTYVAQSELL DAQGNILKG IGFSAPKKQY SLEGDKVEVR LSAPETRGLK  
 151 IDKVYTFETG SYLVNVREDI ANGSGQTANL SADYRIVRDH SEPEGQGYFT  
 201 HSYVGPVYVT PEGNFQKVSF SLDLDDAKSG KSEAEYIRKT PTGWLGMIEH  
 251 HFMSTWILQP KGRQSVCAAG ECNIDIKRRN DKLYSTSVSV PLAAIQNGAK  
 301 AEASINLYAG PQTTSVIANI ADNLQAKDY GKVHWFASPL FWLLNQLHNI  
 351 IGNWGWAIIV LTIIIVKAVLY PLTNASYRSM AKMRAAPKL QAIKEKYGDD  
 401 RMAQQQAMMQ LYTDEKINPL GGCLPMLLQI PVFIGLYWAL FASVELRQAP  
 451 WLGWITDLSR ADPYIILPII MAATMFAQTY LNPPPTDPMQ AKMMKIMPLV  
 25 501 FSVMFFFFPA GLVLYWVNN LLTIAQQWHI NRSIEKQRAQ GEVVS\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a 60kDa inner-membrane protein (accession P25754) of *Pseudomonas putida*

ORF11 and the 60kDa protein show 58% aa identity in 229 aa overlap (BLASTp).

30 ORF11 2 LYAGPQTTSVIANIADNLQAKDYGVHWFASPLFWLLNQLHNIIGNWGWAIIVLTIIIVK 61  
 LYAGP+ S + ++ L+L DYG + + A P+FWLL +H+++GNWGW+IIVLT+++K  
 60K 324 LYAGPKIQSKLKELSPGLELTVDYGFLWFIAQPIFWLLQHIHSLGNWGSIIIVLTMLIK 383  
 ORF11 62 AVLYPLTNASYRSMAMRAAPKLQAIKEKYGDDRRXXXXXXXXXXLYTDEKINPLGGCLPM 121  
 + +PL+ ASYRSM+MRA APKL A+KE++GDDR LY EGINPLGGCLP+  
 35 60K 384 GLFFPLSAASYRSMARMRAVAPKLAALKERFGDDRQKMSQAMMELYKKEKINPLGGCLPI 443  
 ORF11 122 LLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYIILPII MAATMFAQTYLNPPPT 181  
 L+Q+PVF+ LYW L SVE+RQAPW+ WITDLS DP++ILPIIM ATMF Q LNP P  
 40 60K 444 LVQMPVFLALYWVLESVEMRQAPWILWITDLSIKDPFFILPIIMGATMFIQRLNPTTP 503  
 ORF11 182 DPMQAKMMKIMPLVXXXXXXXXXPAGXVLYWVNNLLTIAQQWHINRSIE 230  
 DPMQAK+MK+MP++ PAG VLYWVNN L+I+QQW+I R IE  
 60K 504 DPMQAKVMKMPIIFTFFFLWFPAGLVLYWVNNCLSSISQQWYITRRIE 552

45 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF11 shows 97.9% identity over a 240aa overlap with an ORF (ORF11a) from strain A of *N.*

*meningitidis*:

50 orf11.pep NLYAGPQTTSVIANIADNLQAKDYGVHWF  
 orf11a IKRRNDKLYSTSVSVPLAAIQNGAKSXASINLYAGPQTTSVIANIADNLQAKDYGVHWF  
 280 290 300 310 320 330  
 55 orf11.pep FASPLFWLLNQLHNIIGNWGWAIIVLTIIIVKAVLYPLTNASYRSMAMRAAPKLQAIKE  
 orf11a FASPLFWLLNQLHNIIGNWGWAIIVLTIIIVKAVLYPLTNASYRSMAMRAAPKLQAIKE  
 340 350 360 370 380 390

-87-

		100	110	120	130	140	150
5	orf11.pep	KYGDDRMAQQQAMQLYTDEKINPLGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGI					
	orf11a	KYGDDRMAQQQAMQLYTDEKINPLGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGI					
		400	410	420	430	440	450
10	orf11.pep	TDLSRADPYIILPIIIMAAATMFAQTYLNPPTDPMQAKMMKIMPLVFSXXFFFPAGXVLY					
	orf11a	TDLSRADPYIILPIIIMAAATMFAQTYLNPPTDPMQAKMMKIMPLVFSXXFFFPAGLVLY					
		460	470	480	490	500	510
15	orf11.pep	WVNNLLTIAQQWHINRSIEKQRAQGEVVSX					
	orf11a	WVNNLLTIAQQWHINRSIEKQRAQGEVVSX					
		520	530	540			

The complete length ORF11a nucleotide sequence <SEQ ID 53> is:

20	1	ANGGATTTTA	AAAGACTCAC	NGNGTTTTTC	GCCATCGCAC	TGGTGATTAT
	51	GATCGGATNG	NAAANGATGT	TCCCCACTCC	GAAGCCCGTC	CCCGCGCCCC
	101	AACAGACGGC	ACAACAACAG	GCCGTAANCG	CTTCCGCCGA	AGCCGCGCTC
	151	GCGCCCGNAN	CGCCGATTAC	CGTAACGACC	GACACGGTTC	AAGCCGTCAT
	201	TGATGAAAAA	AGCGGCGACC	TGCGCCGGCT	GACCCTGCTC	AAATACAAAG
25	251	CAACCGGCGA	CNAAAATAAA	CCGTTTCATCC	TGTTTGCGCA	CGGCAANAA
	301	TACACCTACN	TCGCCCANTC	CGAACTTTTG	GACGCGCAGG	GCAACAACAT
	351	TCTAAAAGGC	ATCGGCTTTA	GCGCACCAGAA	AAAACAGTAC	AGCTTGGGAA
	401	GCGACAAAGT	TGAAGTCCGC	CTGAGCGCAC	CTGAAACACG	CGGTCTGAAA
	451	ATCGACAAAG	TTTATACTTT	CACCAAAGGC	AGCTATCTGG	TCAACGTCGG
30	501	CTTCGACATC	GCCAACGGCA	GCGGTCAAAC	CGCCAACCTG	AGCGCGGACT
	551	ACCGCATCGT	CCGCGACCAC	AGCGAACCCG	AGGGTCAAGG	CTACTTTACC
	601	CACTCTTACG	TCGGCCCTGT	TGTTTATACC	CCTGAAGGCA	ACTTCCAAAA
	651	AGTAGCTTTC	TCCGACTTGG	ACGACGATGC	CAANTCCGGN	AAATCCGAGG
	701	CCGAATACAT	CCGCAAAACC	CNGACCGGCT	GGCTCGGCAT	GATTGAACAC
35	751	CACTTCATGT	CCACCTGGAT	CCTCCAACCC	AAAGGCGGAC	AAAGCGTTTG
	801	CGCCGCTGGC	GACTGCNGTA	TNGACATCAA	ACGCCGCAAC	GACAAGCTGT
	851	ACAGCACCAG	CGTCAGCGTG	CCTTTAGCCG	CTATCCAAAA	CGGTGCGAAA
	901	TCNAAGCCT	CCATCAACCT	CTACGCCGGC	CCACAGACCA	CATCNGTTAT
	951	CGCAAAATC	GCCGACAACC	TGCAACTGGN	CAAAGACTAC	GGCAAAGTAC
40	1001	ACTGGTTTCG	CTCCCCCTC	TTTGGCTTTT	TGAACCAACT	GCACAACATC
	1051	ATCGGCAACT	GGGGCTGGGC	GATTATCGTT	TTAACCATCA	TCGTCAAAGC
	1101	CGTACTGTAT	CCATTGACCA	ACGCCTCTTA	CCGTTTCGATG	GCGAAAATGC
	1151	GTGCCGCCGC	GCCCAAACTG	CAAGCCATCA	AAGAGAAATA	CGGCGACGAC
	1201	CGTATGGCGC	AGCAACAAGC	CATGATGCAG	CTTTACACAG	ACGAGAAAAT
45	1251	CAACCCGCTG	GGCGGCTGCC	TGCCATATGCT	GTTGCAAAATC	CCCGTCTTCA
	1301	TCGGATTGTA	TGGGCATTG	TTCGCCTCCG	TAGAATTGCG	CCAGGCACCT
	1351	TGGCTGGGTT	GGATTACCGA	CCTCAGCCGC	GCCGACCNT	ACTACATCCT
	1401	GCCCATCATT	ATGGCGGCAA	CGATGTTTCG	CCAAACCTAT	CTGAACCCGC
	1451	CGCCGACCGA	CCCGATGCAG	GCGAAAATGA	TGAAAATCAT	GCCTTTGGTT
50	1501	NTNTCNNNNA	NGTCTCTCNN	CTTCCCTGCC	GGTCTGGTAT	TGTACTGGGT
	1551	GATCAACAAC	CTCCTGACCA	TCGCCCAGCA	ATGGCACATC	AACCGCAGCA
	1601	TCGAAAAACA	ACGCGCCCAA	GGCGAAGTCG	TTTCTCTAA	

This encodes a protein having amino acid sequence <SEQ ID 54>:

	1	XDFKRLTXFF	AIALVIMIGX	XXMFPTPKPV	PAPQOTAQQQ	AVXASAEAAAL
55	51	APXXPITVTT	DTVQAVIDEK	SGDLRRLTLL	KYKATGDXNK	PFILFGDGKX
	101	YTYXAXSELL	DAQGNNILKG	IGFSAPKKQY	SLEGDKEVR	LSAPETRGLK
	151	IDKVYFTFTK	SYLVNVRFDI	ANGSGQTANL	SADYRIVRDH	SEPEGQGYFT
	201	HSYVGPVYTT	PEGNFQKVSF	SDLDDAXSG	KSEAERYKRT	XTGWLGMIEH
	251	HFMSTWILQP	KGGQSVCAAG	DCXXDIKRRN	DKLYSTSVSV	PLAAIQNGAK
60	301	SXASINLYAG	PQTTSVIANI	ADNLQLXKDY	GKVHWFASPL	FWLLNQLHNI
	351	IGNWGWAIIV	LTIIIVKAVLY	PLTNASYRSM	AKMRAAPKL	QAIKEYGDD
	401	RMAQQQAMMQ	LYTDEKINPL	GGCLPMLLQI	PVFIGLYWAL	FASVELRQAP
	451	WLGWITDLSR	ADPYIILPII	MAATMFAQTY	LNPPTDPMQ	AKMMKIMPLV
	501	XSXXFFXFFA	GLVLYWVINN	LLTIAQQWHI	NRSIEKQRAQ	GEVVS*

ORF11a and ORF11-1 show 95.2% identity in 544 aa overlap:

5	orf11a.pep	XDFKRLTXFFAIALVIMIGXXMFPTPKVPAPQQTAAQQQAVXASAEALAPXXPITVTT
	orf11-1	MDFKRLTAFFAIALVIMIGWEKMFPTPKVPAPQQAQQQAVTASAEALAPATPITVTT
10	orf11a.pep	DTVQAVIDEKSGDLRRLTLLKYKATGDXNKPFILFGDGKXYTYXAXSELLDAQNNILKG
	orf11-1	DTVQAVIDEKSGDLRRLTLLKYKATGDENKPFILFGDGKEYTYVAQSELLDAQNNILKG
15	orf11a.pep	IGFSAPKKQYSLEGDKVEVRLSAPETRGLKIDKVYTFKGSYLVNVRFDIANGSGQTANL
	orf11-1	IGFSAPKKQYSLEGDKVEVRLSAPETRGLKIDKVYTFKGSYLVNVRFDIANGSGQTANL
20	orf11a.pep	SADYRIVRDHSEPEGQGYFTHSYVGPVYTPEGNFQKVSFSDLLDDAKSGKSEAERYIKRT
	orf11-1	SADYRIVRDHSEPEGQGYFTHSYVGPVYTPEGNFQKVSFSDLLDDAKSGKSEAERYIKRT
25	orf11a.pep	XTGWLGMIEHHFMSTWILQPKGQSVCAAGDCXXDIKRRNDKLYSTSVSVPLAAIQNGAK
	orf11-1	PTGWLGMIEHHFMSTWILQPKGRQSVCAAGECNIDIKRRNDKLYSTSVSVPLAAIQNGAK
30	orf11a.pep	SKASINLYAGPQTTSVIANIADNLQKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIV
	orf11-1	AEASINLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIV
35	orf11a.pep	LTIIVKAVLYPLTNASYRSMAKMRAAAPKLQAIKEKYGDDRMAQQQAMMQLYTDEKINPL
	orf11-1	LTIIVKAVLYPLTNASYRSMAKMRAAAPKLQAIKEKYGDDRMAQQQAMMQLYTDEKINPL
40	orf11a.pep	GGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIIMAAATMFAQTY
	orf11-1	GGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIIMAAATMFAQTY
45	orf11a.pep	LNPPPTDEMQAKMMKIMPLVXSXXFFXFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQ
	orf11-1	LNPPPTDEMQAKMMKIMPLVFSVMFFFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQ
50	orf11a.pep	GEVVSX
	orf11-1	GEVVSX

# 60 Homology with a predicted ORF from *N.gonorrhoeae*

ORF11 shows 96.3% identity over a 240aa overlap with a predicted ORF (ORF11.ng) from *N. gonorrhoeae*:

65	Orf11	NLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLT	57
	orf11ng	MAVNLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLT	60



	orf11	IIVKAVLYPLTNASYRSMAKMRAAAPKLQAIKEKYGDDRMAQQQAMMQLYTDEKINPLGG	117
	orf11ng	IIVKAVLYPLTNASYRSMAKMRAAAPELQTIKEKYGDDRMAQQQAMMQLFEDEEINPLGG	120
5	orf11	CLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTYLN	177
	orf11ng	CLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTYLN	180
10	orf11	PPPTDPMQAKMMKIMPLVFSXFFFFFPAGXVLYWVNNLLTIAQQWHINRSIEKQRAQGE	237
	orf11ng	PPPTDPMQAKMMKIMPLVFSVMFFFFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQGE	240
15	orf11	VVS 240	
	orf11ng	VVS 243	

An ORF11ng nucleotide sequence <SEQ ID 55> was predicted to encode a protein having amino acid sequence <SEQ ID 56>:

20	1	MAVNLYAGPQ	TTSVIANIAD	NLQLAKDYGK	VHWFASPLEW	LLNQLHNIIG
	51	NQGWAIIVLT	IIVKAVLYPL	TNASYRSMK	MRAAAPELQT	IKEKYGDDRM
	101	AQQQAMMQLF	EDEEINPLGG	CLPMLLQIPV	FIGLYWALFA	SVELRQAPWL
	151	GWITDLSRAD	PYYILPIIMA	ATMFAQTYLN	PPPTDPMQAK	MMKIMPLVFS
	201	VMFFFFPAGL	VLYWVNNLL	TIAQQWHINR	SIEKQRAQGE	VVS*

Further sequence analysis revealed the complete gonococcal DNA sequence <SEQ ID 57> to be:

25	1	ATGGATTTTA	AAAGACTCAC	GGCGTTTTTC	GCCATCGCGC	TGGTGATTAT
	51	GATCGGCTGG	GAAAAAATGT	TCCCCACCCC	GAAACCCGTC	CCCGCGCCCC
	101	AACAGGCGGC	ACAAAAACAG	GCAGCAACCG	CTTCCGCGCA	AGCCGCGCTC
	151	GCGCCGCGCA	CGCCGATTAC	CGTAACGACC	GACACGGTTC	AAGCCGTAT
30	201	TGATGAAAAA	AGTGGCGACC	TGCGCCGGCT	GACCCTGCTC	AAATACAAAG
	251	CAACCGGCGA	CGAAAAACAA	CCGTTCGTCC	TGTTTGCGCA	CGGCAAAGAA
	301	TACACCTACG	TCGCCCAATC	CGAACTTTTG	GACGCGCAGG	GCAACAACAT
	351	TCTGAAAGGC	ATCGGCTTTA	GCGCACCGAA	AAAACAGTAC	ACCCTCAACG
35	401	GCGACACAGT	CGAAGTCCGC	CTGAGCGCGC	CCGAAACCAA	CGGACTGAAA
	451	ATCGACAAAG	TCTATACCTT	TACCAAAGAC	AGCTATCTGG	TCAACGTCCG
	501	CTTCGACATC	GCCAACGGCA	GCGGTCAAAC	CGCCAACCTG	AGCGCGGACT
	551	ACCGCATCGT	CCGCGACCAC	AGCGAACCAG	AGGGTCAAGG	CTACTTTACC
40	601	CACCTCTACG	TCGGCCCTGT	TGTTTATACC	CCTGAAGGCA	ACTTCCAAAA
	651	AGTCAGCTTC	TCCgacTTgg	acgACGATGC	gaaaTccggc	aaATccgagg
	701	ccgaatacaT	CCGCAAAACC	ccgaccggtt	ggctcggcat	gattgaacac
	751	cacttcatgt	ccacctggat	cctccAAcct	aaagggcgcc	aaaacgtttg
45	801	cgcccaggga	gactgcccgt	tcgacattaa	aCgcccgaac	gacaagctgt
	851	acagcgcaag	cgtcagcgtg	cctttaaccg	ctatcccaac	ccggggggcca
	901	aaaccgaaaa	tggcggTCAA	CCTGTATGCC	GGTCCGCAAA	CCACATCCGT
	951	TATCGCAAAC	ATCGCcgacA	ACCTGCAACT	GGCAAAAGAC	TACGGTAAAG
50	1001	TACTACTGTT	CGCATGCGCG	CTCTTCTGGC	TCCTGAACCA	ACTGCACAAC
	1051	ATTATCGGCA	ACTGGGGCTG	GGCAATCGTC	GTTTTGACCA	TCATCGTCAA
	1101	AGCCGTACTG	TATCCATTGA	CCAACGcttc	ctACCGTTTC	ATGGCGAAAA
	1151	TGCGTGccgc	cgcacCcaaA	CTGCAGACCA	TCAAAGAAAA	ATAcgGCGAC
55	1201	GACCGTATGG	CGCAACAGCA	AGCGATGATG	CAGCTTTACA	AAGacgAGAA
	1251	AATCAACCCG	CTGGGCGGCT	GTctgcctat	gctgttgCAA	ATCCCCGTCT
	1301	TCATCGGCTT	GTACTGGGCA	TTGTTTCGCT	CCGTAGAATT	GCGCCAGGCA
	1351	CCTTGGCTGG	GCTGGATTAC	CGACCTCAGC	CGCGCCGACC	CCTACTACAT
	1401	CCTGCCCATC	ATTATGGCGG	CAACGATGTT	CGCCCAAAACC	TATCTGAACC
	1451	CGCCGCCGAC	CGACCCGATG	CAGGCGAAAA	TGATGAAAAT	CATGCCGTTG
55	1501	GTTTCTCCG	TCATGTTCTT	CTTCTTCCCT	GCCGGTTTGG	TTCTCTACTG
	1551	GGTGGTCAAC	AACCTCCTGA	CCATCGCCCA	GCAGTGGCAC	ATCAACCGCA
	1601	GCATCGAAAA	ACAACGCGCC	CAAGCGGAAG	TCGTTTCCTA	A

This encodes a protein having amino acid sequence <SEQ ID 58; ORF11ng-1>:

60	1	MDFKRLTAFF	AIALVIMIGW	EKMFPTEPKPV	PAPOQAAQKQ	AATASAEAL
	51	APATPITVTT	DTVQAVIDEK	SGDLRRLTL	KYKATGDENK	PFVLFGDGKE
	101	ITYVAQSELL	DAQGNNILKG	IGFSAPKKQY	TLNGDTVEVR	LSAPETNGLK
	151	IDKVYTFTKD	SYLVNVRFDI	ANGSGQTANL	SADYRIVRDH	SEPEGQGYFT
	201	HSYVGPVYVT	PEGNFQKVSF	SDLDDAKSG	KSEAEYIRKT	PTGWLGMIEH
	251	HFMSTWILQP	KGQNVCAQG	DCRIDIKRRN	DKLYSASVSV	PLTAIPTRGP
	301	KPKMAVNLYA	GPQTTSVIAN	IADNLQLAKD	YGKVHWFASF	LFWLLNQLHN

```

351 IIGNWGWAIV VLTIIIVKAVL YPLTNASYRS MAKMRAAPK LQTIKEYGD
401 DRMAQQQAMM QLYKDEKINP LGGCLPMLLQ IPVFIGLYWA LFASVELRQA
451 PWLGWITDLS RADPYYILPI IMAATMFAQT YLNPPPTDPM QAKMMKIMPL
501 VFSVMFFFEF AGLVLYWVWN NLLTIAQQWH INRSIEKQRA QGEVVS*

```

5 ORF11ng-1 and ORF11-1 shown 95.1% identity in 546 aa overlap:

```

10 orf11ng-1.pep 10 20 30 40 50 60
    MDFKRLTAFFAIALVIMIGWEKMFPTPKPVPAPQQAQQAATASAEALAPATPITVTT
    orf11-1      10 20 30 40 50 60
    MDFKRLTAFFAIALVIMIGWEKMFPTPKPVPAPQQAQQAATASAEALAPATPITVTT

15 orf11ng-1.pep 70 80 90 100 110 120
    DTVQAVIDEKSGDLRRLTLTKYKATGDENKPFVLFGDGKEYTYVAQSELLDAQGNILKG
    orf11-1      70 80 90 100 110 120
    DTVQAVIDEKSGDLRRLTLTKYKATGDENKPFILFGDGKEYTYVAQSELLDAQGNILKG

20 orf11ng-1.pep 130 140 150 160 170 180
    IGFSAPKKQYTLNGDTVEVRLSAPETNGLKIDKVYFTFKDSYLVNVREDIANGSGQTANL
    orf11-1      130 140 150 160 170 180
    IGFSAPKKQYTLNGDTVEVRLSAPETNGLKIDKVYFTFKDSYLVNVREDIANGSGQTANL

25 orf11ng-1.pep 190 200 210 220 230 240
    SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDDAKSGKSEAEYIRKT
    orf11-1      190 200 210 220 230 240
    SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDDAKSGKSEAEYIRKT

30 orf11ng-1.pep 250 260 270 280 290 300
    PTGWLGMIEHHFMSTWILQPKGQNVCAQGD CRIDIKRRNDKLYSASVSVPLTAIPTRGP
    orf11-1      250 260 270 280 290
    PTGWLGMIEHHFMSTWILQPKGRQSVCAAGECNIDIKRRNDKLYSTSVSVPLAAIQN-GA

35 orf11ng-1.pep 310 320 330 340 350 360
    KPKMAVNLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIV
    orf11-1      300 310 320 330 340 350
    KAEASINLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAII

40 orf11ng-1.pep 370 380 390 400 410 420
    VLTIIIVKAVLYPLTNASYRSMAKMRAAPKLQTIKEYGDDRMAQQQAMMQLYKDEKINP
    orf11-1      360 370 380 390 400 410
    VLTIIIVKAVLYPLTNASYRSMAKMRAAPKLQAIKEYGDDRMAQQQAMMQLYTDEKINP

45 orf11ng-1.pep 430 440 450 460 470 480
    LGGCLEPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQT
    orf11-1      420 430 440 450 460 470
    LGGCLEPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQT

50 orf11ng-1.pep 490 500 510 520 530 540
    YLNPPPTDPMQAKMMKIMPLVFSVMFFFPAGLVLYWVWNLLTIAQQWHINRSIEKQRA
    orf11-1      480 490 500 510 520 530
    YLNPPPTDPMQAKMMKIMPLVFSVMFFFPAGLVLYWVWNLLTIAQQWHINRSIEKQRA

60 orf11ng-1.pep QGEVVSX
    orf11-1      QGEVVSX
    540

```

65 In addition, ORF11ng-1 shows significant homology with an inner-membrane protein from the database (accession number p25754):

-91-

ID 60IM PSEPU STANDARD; PRT; 560 AA.  
 AC P25754;  
 DT 01-MAY-1992 (REL. 22, CREATED)  
 DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE 60 KD INNER-MEMBRANE PROTEIN. . . .

SCORES Initl: 1074 Initn: 1293 Opt: 1103  
 Smith-Waterman score: 1406; 41.5% identity in 574 aa overlap

		10	20	30	40
orf11ng-1.pep	MDFKR---	LTAF	FAIALVIMIGW----	EKM	FPT-----PKPV
	:	: :	: :	: :	: :
p25754	MDIKRTILIAALAVSVYVMVLKWNDDYGQAALPTQNTAASTVAPGLPDGVPAGNNGASAD				
		10	20	30	40
		50	60	70	80
orf11ng-1.pep	AATASAEALAPATPIT-----	VTTD	TVQAVIDEKSGDLRRLTLLKYKATGDE-NKPF		
	: :	: :	: :	: :	: :
p25754	VPSANAESSPAELAPVALSKDLIRVKTDVLELAIDPVGGDIVQLNLPKYPRRQHPNIPF				
		70	80	90	100
		110	120	130	140
orf11ng-1.pep	VLFGDGKEYTYVAQSELLDAQNNILKGIG---	FSAP	KKQYTL-NGD---TVEVRLSAPE		
	:	: :	: :	: :	: :
p25754	QLFDNGGERVYLAQSGLTGTDGPDA-RASGRPLYAAEQKSYQLADGQEQLVVDLKFS---				
		130	140	150	160
		170	180	190	200
orf11ng-1.pep	TNGLKIDKVYTFTKDSYLVNVRFDIANGSGQTANLSADYRIVRDHS-EPEGQGYF-THSY				
	: :	: :	: :	: :	: :
p25754	DNGVNYIKRFSFKRGEYDLNVSYLIDNQSGQAWNGNMFAQLKRDASGDPSSSTATGTATY				
		180	190	200	210
		220	230	240	250
orf11ng-1.pep	VGPVVYTPGPNFQKVSFSDLDLDDAKSGKSEAERYIKRTPGTGWLGMIEHHFMSTWILQPKGG				
	: :	: :	: :	: :	: :
p25754	LGAALWTASEPYKKVSMKDID---KGSLE-----NVSGGWVWALQHYFVTAWI-PAKSD				
		240	250	260	270
		280	290	300	310
orf11ng-1.pep	QNVCAQGDCCRIDIKRRNDKLYSASVSVPLTAIPTRGPKPKMAVNLYAGPQTTSVIANIAD				
	: :	: :	: :	: :	: :
p25754	NNV-----VQTRKDSQGNYYIGYTGPIVISVPA-GGKVETSALLYAGPKIQSKLKELSP				
		290	300	310	320
		330	340	350	360
orf11ng-1.pep	NLQLAKDYGVKWHWF-ASPLFWLLNQLHNIIGNWGWAIIVLTIIVKAVLYPLTNASYRSMA				
	: :	: :	: :	: :	: :
p25754	GLELTVDYGF-LWFIAQPIFWLLQHIHSLGNWGSIIIVLTMLIKGLFFPLSAASYRSMA				
		340	350	360	370
		380	390	400	410
orf11ng-1.pep	KMRAAAPKLQTIKEKYGDDRMAQQQAMMQLYKDEKINPLGGCLPMLLQIPVFIGLYWALF				
	: :	: :	: :	: :	: :
p25754	RMRAVAPKLAALKERFGDDRQKMSQAMMELYKKEKINPLGGCLPILVQMPVFLALYVWLL				
		400	410	420	430
		440	450	460	470
orf11ng-1.pep	ASVELRQAPWLGWITDLSRADPYIILPIIMAATMFAQTYLNPPPTDPMQAKMMKIMPLVF				
	: :	: :	: :	: :	: :
p25754	ESVEMRQAPWILWITDLSIKDPFFILPIIMGATMFIQQLNPTPPDPMQAKVMKMMPIIF				
		460	470	480	490
		500	510	520	530
orf11ng-1.pep	SVMFFFFFFAGLVLYWVNNLLTIAQQWHINRSIEKQRAQGEVVSX				
	: :	: :	: :	: :	: :
p25754	TFFFLWFPAGLVLYWVNNCLSSQQWYITRRIEATKKAAA				
		520	530	540	550
		560			

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 59>:

15 This corresponds to the amino acid sequence <SEQ ID 60; ORF13>:

Further sequence analysis elaborated the DNA sequence slightly <SEQ ID 61>:

This corresponds to the amino acid sequence <SEQ ID 62; ORF13-1>:

Computer analysis of this amino acid sequence gave the following results:

ORF13 shows 92.9% identity over a 126aa overlap with an ORF (ORF13a) from strain A of *N.*

```

                                10      20      30      40      50
orf13.pep      AVLIIELLTGTVYLLVVSAALAGSGIAYGLTGSTPAAVLTXALLSALGIXF
                |||
orf13a      MTVWFVAAVAVLIIELLTGTVYLLVVSAALAGSGIAYGLTGSTPAAVLTAALLSALGIWF
                10      20      30      40      50
                                60      70      80      90      100     110
orf13.pep      VHAKTAVRKVETDSYQDLDAGQYVEILRHTGGNRYEVXYRGTXWQAQNTGQEELEPGTRA
                |||
orf13a      VHAKTAVGKVETDSYQDLDAGQYAEILRHAGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
                70      80      90      100     110     120
                                120
orf13.pep      LIVRKEGNLLIITHPX
                |||

```

orf13a LIVRKEGNLLIIAKPX  
130

The complete length ORF13a nucleotide sequence <SEQ ID 63> is:

```

1  ATGACTGTAT GGTGTGTTGC CGCTGTTGCC GTCTTAATCA TCGAATTATT
5  51  GACGGGAACG GTTTATCTTT TGGTTGTCAG CGCGGCTTTG GCGGGTTCGG
101 GCATTGCTTA CCGGCTGACC GGCAGCACGC CTGCCGCCGT CTTGACCGCC
151 GCTCTGCTTT CCGGCTGGG TATTTGGTTC GTACACGCCA AAACCGCCGT
201 GGGAAAAGTT GAAACGGATT CATATCAGGA TTTGGATGCC GGGCAATATG
251 CCGAAATCCT CCGGCACGCA GCGGCAACC GTTACGAAGT TTTTATCGC
10 301 GGTACGCACT GGCAGGCTCA AAATACGGGG CAAGAAGAGC TTGAACCAGG
351 AACGCGCGCC CTAATCGTCC GCAAGGAAGG CAACCTTCTT ATCATCGCAA
401 AACCTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 64>:

```

1  MTVWFVAAVA VLIIELLTGT VYLLVVSAA AGSGIAYGLT GSTPAAVLTA
15 51  ALLSALGIWF VHAKTAVGKV ETDSYQDLDA GQYAEILRHA GGNRYEVFYR
101 GTHWQAQNTG QEELEPGTRA LIVRKEGNLL IIAKP*

```

ORF13a and ORF13-1 show 94.4% identity in 126 aa overlap

```

20 orf13a.pep      10      20      30      40      50      60
    MTVWFVAAVAVLIIELLTGTVYLLVVSAAAGSGIAYGLTGSTPAAVLTAALLSALGIWF
orf13-1          10      20      30      40      50
    AVLIIELLTGTVYLLVVSAAAGSGIAYGLTGSTPAAVLTXALLSALGIXF

25 orf13a.pep      70      80      90      100     110     120
    VHAKTAVGKVETDSYQDLDAQYAEILRHAGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf13-1          70      80      90      100     110     120
    VHAKTAVRKVETDSYQDLDAQYVEILRHTGGNRYEVFYRGTHWQAQNTGQEELEPGTRA

30 orf13a.pep      130
    LIVRKEGNLLIIAKPX
orf13-1          120
    LIVRKEGNLLIITHPX

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF13 shows 89.7% identity over a 126aa overlap with a predicted ORF (ORF13.ng) from *N.gonorrhoeae*:

```

40 orf13          AVLIIELLTGTVYLLVVSAAAGSGIAYGLTGSTPAAVLTXALLSALGIXF 51
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf13ng      MTVWFVAAVAVLIIELLTGTVYLLVVSAAAGSGIAYGLTGSTPAAVLTAALLSALGIWF 60

orf13          VHAKTAVRKVETDSYQDLDAQYVEILRHTGGNRYEVXYRGTXWQAQNTGQEELEPGTRA 111
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
45 orf13ng      VHAKTAVGKVETDSYQDLDTGKYAEILRYTGGNRYEVFYRGTHWQAQNTGQEVFEPGTRA 120

orf13          LIVRKEGNLLIITHP 126
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf13ng      LIVRKEGNLLIIANP 135

```

50 The complete length ORF13ng nucleotide sequence <SEQ ID 65> is:

```

1  ATGACTGTAT GGTGTGTTGC CGCTGTTGCC GTCTTAATCA TCGAATTATT
51  GACGGGAACG GTTTATCTTT TGGTTGTCAG CGCGGCTTTG GCGGGTTCGG
101 GCATTGCCTA CCGGCTGACT GGCAGCACGC CTGCCGCCGT CTTGACCGCC
151 GCACTGCTTT CCGGCTGGG CATTTGGTTC GTACATGCCA AAACCGCCGT
55 201 GGGAAAAGTT GAAACGGATT CATATCAGGA TTTGGATACC GGAAATATG
251 CCGAAATCCT CCGATACACA GCGGCAACC GTTACGAAGT TTTTATCGC
301 GGTACGCACT GGCAGGCGCA AAATACGGGG CAGGAAGTGT TTGAACCGGG
351 AACGCGCGCC CTCATCGTCC GCAAAGAAGG TAACCTTCTT ATCATCGCAA
401 ACCCTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 66>:

```

1  MTVWFVAAVA VLIIELLTGT VYLLVVSAA AGSGIAYGLT GSTPAAVLTA
51  ALLSALGIWF VHAKTAVGKV ETDSYQDLDT GKAEILRYT GGNRYEVFYZ
101 GTHWQAQNTG QEVFEPGTRA LIVRKEGNLL IIANP*

```

5 ORF13ng shows 91.3% identity in 126 aa overlap with ORF13-1:

```

10  orf13-1.pep      10      20      30      40      50
      AVLIIELLTGT VYLLVVSAA AGSGIAYGLT GSTPAAVLTXALLSALGIXF
orf13ng      MTVWFVAAVA VLIIELLTGT VYLLVVSAA AGSGIAYGLT GSTPAAVLTAALLSALGIWF
      10      20      30      40      50      60

15  orf13-1.pep      60      70      80      90      100     110
      VHAKTAVRKVETDSYQDL DAGQYVEILRHTGGNRYEVFYZRGTHWQAQNTGQEELEPGTRA
orf13ng      VHAKTAVGKVETDSYQDLDTGKYAEILRYTGGNRYEVFYZRGTHWQAQNTGQEVFEPGTRA
      70      80      90      100     110     120

20  orf13-1.pep      120
      LIVRKEGNLLIITHPX
orf13ng      LIVRKEGNLLIIANPX
      130

```

Based on this analysis, including the extensive leader sequence in this protein, it is predicted that

25 ORF13 and ORF13ng are likely to be outer membrane proteins. It is thus predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 9

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 67>:

```

30  1  ATGTWTGATT TCGGTTTGG CGACTGGTT TTTGTCGGCA TTATCGCCCT
51  GATWtCCTC GGCCCCGAAC GCsTGCCCGA GGCCGCCCGC AycGCCGGAC
101 GGcTCATCGG CAGGCTGCAA CGCTTTGTCG GcAGCGTCAA ACAGGAATTT
151 GAACTCAA TCGAACTGGA AGAACTGAGG AAGGCAAAGC AGGAATTTGA
201 AGCTGCCGcC GCTCAGGTTC GAGACAGCCT CAAAGAAACC GGTACGGATA
35  251 TGGAAGGCAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
301 CTGCCCGAAC AGCGGACACC TGCCGATTTC GGTGTCGATG AAAACGGCAA
351 TCCGCT.TCC CGATGCGGCA AACACCCTAT CAGACGGCAT TTCCGACGTT
401 ATGCCGTC..

```

This corresponds to the amino acid sequence <SEQ ID 68; ORF2>:

```

40  1  MXDFGLGELV FVGIIALIVL GPERXPEAR XAGRLIGRLQ RFVGSVKQEF
51  DTQIELEELR KAKQEFEEAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK
101 LPEQRTPADE GVDENGNPXS RCGKHPIRRH FFRYAV..

```

Further work revealed the complete nucleotide sequence <SEQ ID 69>:

```

45  1  ATGTTTGATT TCGGTTTGGG CGAGCTGGTT TTTGTCGGCA TTATCGCCCT
51  GATTGTCCTC GGCCCCGAAC GCCTGCCCGA GGCCGCCCGC ACCGCCGGAC
101 GGCTCATCGG CAGGCTGCAA CGCTTTGTCG GCAGCGTCAA ACAGGAATTT
151 GAACTCAA TCGAACTGGA AGAACTGAGG AAGGCAAAGC AGGAATTTGA
201 AGCTGCCGCC GCTCAGGTTC GAGACAGCCT CAAAGAAACC GGTACGGATA
50  251 TGGAAGGCAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
301 CTGCCCGAAC AGCGGACACC TGCCGATTTC GGTGTCGATG AAAACGGCAA
351 TCCGCTTCCC GATGCGGCAA ACACCCTATC AGACGGCATT TCCGACGTTA
401 TGCCGTCCGA ACGTTCTTAC GCTTCCGCCG AAACCCTTGG GGACAGCGGG
451 CAAACCGGCA GTACAGCCGA ACCCGCGGAA ACCGACCAAG ACCGCGCATG
501 GCGGGAATAC CTGACTGCTT CTGCCGCCGC ACCCGTCGTA CAGACCGTCG

```

551 AAGTCAGCTA TATCGATACT GCTGTTGAAA CGCCTGTTCC GCACACCACT  
 601 TCCCTGCGCA AACAGGCAAT AAGCCGCAAA CGCGATTTTC GTCCGAAACA  
 651 CCGCGCCAAA CCTAAATTGC GCGTCCGTAA ATCATAA

This corresponds to the amino acid sequence <SEQ ID 70; ORF2-1>:

5 1 MFDFGLGELV FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEF  
 51 DTQIELEELR KAKQEFEEAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK  
 101 LPEQRTPADF GVDENGNPLP DAANTLSDGI SDVMPSESY ASAETLGDSG  
 151 QTGSTAEPAE TDQDRAWREY LTASAAAPVV QTVEVSYIDT AVETVPVPHTT  
 201 SLRKQAI SRK RDLRPK HRAK PKLRVRKS\*

10 Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 71 >:

1 ATGTTTGATT TCGGTTTGGG CGAGCTGGTT TTTGTCGGCA TTATCGCCCT  
 51 GATTGTCCTC GGCCCCGAAC GCCTGCCCCG GGCCGCCCGC ACCGCCGGAC  
 101 GGCTCATCGG CAGGCTGCAA CGCTTTGTCG GCAGCGTCAA ACAGGAATTT  
 151 GACACGCAAA TCGAACTGGA AGAACTAAGG AAGGCAAAGC AGGAATTTGA  
 201 AGCTGCCGCT GCTCAGGTTT GAGACAGCCT CAAAGAAACC GGTACGGATA  
 251 TGGAGGGTAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA  
 301 CTGCCCCAAC AGCGCACGCC TGCTGATTTC GGTGTGCGATG AAAACGGCAA  
 351 TCCCTTTCCC GATGCGGCAA ACACCCTATT AGACGGCATT TCCGACGTTA  
 401 TGCCGTCCGA ACGTTCTTAC GCTTCCGCGG AAACCCCTGG GGACAGCGGG  
 451 CAAACCGGCA GTACAGCCGA ACCCGCGGAA ACCGACCAAG ACCGTGCATG  
 501 GCGGGAATAC CTGACTGCTT CTGCCGCCGC ACCCGTCGTA CAGACCGTCG  
 551 AAGTCAGCTA TATCGATACC GCTGTTGAAA CCCCTGTTCC GCATACCACT  
 601 TCGCTGCGTA AACAGGCAAT AAGCCGCAAA CGCGATTTGC GTCCTAAATC  
 651 CCGCGCCAAA CCTAAATTGC GCGTCCGTAA ATCATAA

25 This encodes a protein having amino acid sequence <SEQ ID 72; ORF2a>:

1 MFDFGLGELV FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEF  
 51 DTQIELEELR KAKQEFEEAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK  
 101 LPEQRTPADF GVDENGNPLP DAANTLLDGI SDVMPSESY ASAETLGDSG  
 151 QTGSTAEPAE TDQDRAWREY LTASAAAPVV QTVEVSYIDT AVETVPVPHTT  
 30 201 SLRKQAI SRK RDLRPK HRAK PKLRVRKS\*

The originally-identified partial strain B sequence (ORF2) shows 97.5% identity over a 118aa overlap with ORF2a:

		10	20	30	40	50	60
35	orf2.pep	MXDFGLGELVFVGIIALIVL	GPERXPEAARXAGRLIGRLQRFVGSVKQEF	DTQIELEELR			
	orf2a	MFDFGLGELVFVGIIALIVL	GPERLPEAARTAGRLIGRLQRFVGSVKQEF	DTQIELEELR			
		10	20	30	40	50	60
40	orf2.pep	KAKQEFEEAAAQVRDSLKET	GTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPXS				
	orf2a	KAKQEFEEAAAQVRDSLKET	GTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPFP				
		70	80	90	100	110	120
45	orf2.pep	RCGKHPIRRHFRRYAV					
	orf2a	DAANTLLDGISDVMPSESYASAETLGDSG	QTGSTAEP	PAETDQDRAWREY	LTASAAAPVV		
		130	140	150	160	170	180

50 The complete strain B sequence (ORF2-1) and ORF2a show 98.2% identity in 228 aa overlap:

	orf2a.pep	MFDFGLGELVFVGIIALIVL	GPERLPEAARTAGRLIGRLQRFVGSVKQEF	DTQIELEELR	60
	orf2-1	MFDFGLGELVFVGIIALIVL	GPERLPEAARTAGRLIGRLQRFVGSVKQEF	DTQIELEELR	60
55	orf2a.pep	KAKQEFEEAAAQVRDSLKET	GTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPFP		120
	orf2-1	KAKQEFEEAAAQVRDSLKET	GTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPLP		120
60	orf2a.pep	DAANTLLDGISDVMPSESYASAETLGDSG	QTGSTAEP	PAETDQDRAWREY	LTASAAAPVV 180

orf2-1 DAANTLSDGISDVMPSESYASAETLGDGSGQTGSTAEPAETDQDRAWREYLTASAAAPVV 180  
 orf2a.pep QTVEVSYIDTAVETFPVPHTTSLRKQAISRKRDLRPKSRAKPKLRVRKSX 229  
 |||||:|||||  
 5 orf2-1 QTVEVSYIDTAVETFPVPHTTSLRKQAISRKRDFRPHKRAKPKLRVRKSX 229

Further work identified a partial DNA sequence <SEQ ID 73> in *N.gonorrhoeae* encoding the following amino acid sequence <SEQ ID 74; ORF2ng>:

1 MFDFGLGELI FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEL  
 51 DTQIELEELR KVKQAFEAAA AQVRDSLKET DTDQMNSLHD ISDGLKPWEK  
 101 LPEQRTPADF GVDEKGNLSL RYGKHIRRH FRRYAV\*

Further work identified the complete gonococcal gene sequence <SEQ ID 75>:

1 ATGTTTGATT TCGGTTTGGG CGAGCTGATT TTTGTGCGCA TTATCGCCCT  
 51 GATTGTCCTT GGTCAGAAC GCCTGCCCGA AGCCGCCCGC ACTGCCGGAC  
 101 GGCTTATCGG CAGGCTGCAA CGCTTTGTAG GAAGCGTCAA ACAAGAACTT  
 151 GAACTCAAA TCGAACTGGA AGAGCTGAGG AAGGTCAAGC AGGCATTCTGA  
 201 AGCTGCCGCC GCTCAGGTTT GAGACAGCCT CAAAGAAACC GATACGGATA  
 251 TGCAGAACAG TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA  
 301 CTGCCCCGAA AGCGCACGCC tgcgatttc gGTGTCGATg AAAacggcaa  
 351 tccccctccc gATACGGCAA ACACCGTATC AGACGGCATT TCCGACGTTA  
 401 TGCCGCTCTGA ACCTTCCGAT ACTtccgcCG AAACCCCTTG GGACGACAGG  
 451 CAAACCGGCA GTACAGCCGA ACCTGCGGAA ACCGACAAAG ACCGCGCATG  
 501 GCGGGAATAC CTGactgctt ctgcccgcgc acctgtcgta Cagagggccg  
 551 tcgaagtcag ctaTATCGAT ACTGCTGTTG AAacgcctgT tccgcaCacc  
 601 acttccctgc gcaAACAGGC AATAAACCGC AAACGCGATT TttgtccgaA  
 651 ACACCGCGCc aACCGAAat tgcgcgtcCG TAAATCATAA

This encodes a protein having the amino acid sequence <SEQ ID 76; ORF2ng-1>:

1 MFDFGLGELI FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEL  
 51 DTQIELEELR KVKQAFEAAA AQVRDSLKET DTDQMNSLHD ISDGLKPWEK  
 101 LPEQRTPADF GVDENGNPLP DTANTVSDGI SDVMPSESD TSAETLGDGR  
 151 QTGSTAEPAE TDKDRAWREY LTASAAAPVV QRAVEVSYID TAVETFPVHT  
 201 TSLRKQAINR KRDFCPKHA KPKLRVRKS\*

The originally-identified partial strain B sequence (ORF2) shows 87.5% identity over a 136aa overlap with ORF2ng:

35 orf2.pep MXDFGLGELVFVGIIALIVLGPXPEAARXAGRLIGRLQRFVGSVKQEFDTQIELEELR 60  
 orf2ng MFDFGLGELIFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR 60  
 40 orf2.pep KAKQFEAAAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPXS 120  
 orf2ng KVKQAFEAAAAAQVRDSLKETDTDMQNSLHDISDGLKPWEKLPEQRTPADFGVDENGNPLP 120  
 orf2.pep RCGKHPIRRHFRRYAV 136  
 45 orf2ng RYGKHIRRHFRYAV 136

The complete strain B and gonococcal sequences (ORF2-1 & ORF2ng-1) show 91.7% identity in 229 aa overlap:

50 orf2-1.pep MFDFGLGELVFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR  
 orf2ng-1 MFDFGLGELIFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR  
 10 20 30 40 50 60  
 55 orf2-1.pep KAKQFEAAAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPLP  
 orf2ng-1 KVKQAFEAAAAAQVRDSLKETDTDMQNSLHDISDGLKPWEKLPEQRTPADFGVDENGNPLP  
 70 80 90 100 110 120



-97-

		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf2-1.pep	DAANTLS	DGISD	MPSE	RSYAS	ETLGD	SGQTG
	orf2ng-1	DTANTV	SDGIS	DMPSE	RSYAS	ETLGD	SGQTG
		130	140	150	160	170	180
10	orf2-1.pep	Q-TVEV	SYIDT	AVETP	VPHTT	SLRKQ	AI
	orf2ng-1	QRAVEV	SYIDT	AVETP	VPHTT	SLRKQ	AI
		190	200	210	220	229	
		Q-TVEV	SYIDT	AVETP	VPHTT	SLRKQ	AI
		190	200	210	220	230	

Computer analysis of these amino acid sequences indicates a transmembrane region (underlined),  
and also revealed homology (59% identity) between the gonococcal sequence and the TatB protein  
of *E.coli*:

gnl|PID|e1292181 (AJ005830) TatB protein [Escherichia coli] Length = 171  
Score = 56.6 bits (134), Expect = 1e-07  
Identities = 30/88 (34%), Positives = 52/88 (59%), Gaps = 1/88 (1%)  
Query: 1 MFD FGLGELIFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR 60  
Sbjct: 1 MFDIGFSELLLVFIIGLVLPQRLPVAVKTVAGWIRALRSLATTVQNELTQELKLQEFQ 60  
Query: 61 -KVKQAFEAAAAQVRDSLKETDTDMQNS 87  
Sbjct: 61 DSLKKVEKASLTNLTPELKASMDLRQA 88

Based on this analysis, it was predicted that ORF2, ORF2a and ORF2ng are likely to be membrane  
proteins and so the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be  
useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF2-1 (16kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above.  
The products of protein expression and purification were analyzed by SDS-PAGE. Figure 3A  
shows the results of affinity purification of the GST-fusion protein, and Figure 3B shows the results  
of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice,  
whose sera were used for Western blots (Figure 3C), ELISA (positive result), and FACS analysis  
(Figure 3D). These experiments confirm that ORF37-1 is a surface-exposed protein, and that it is  
a useful immunogen.

### Example 10

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 77>:

1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC  
51 CGC.TGCGGG AACTGACAG GTATTCCATC GCATGGCGGA GkTAAACgCT  
101 TTgCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTTAAA  
151 GACATGGATT TACAGGCATT ACACGGACGA AAGATTGCAT TGTACATTGC  
201 CACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGG TCGCTACTCC  
251 ATTGATGCAC kGrTwCstGG CGAATACATA AACAGCCCTG CCGTCCGTAC  
301 CGATTACACC TATCCACGTT ACGAAACCAC CGCTGAAACA ACATCAGGCG  
351 GTTTGACAGG TTTAACCCTT TCTTTATCTA CACTTAATGC CCCTGCACTC  
401 TCTCGCACCC AATCAGACGG TAGCGGAAGT AAAAGCAGTC TGGGCTTAAA  
451 TATTGGCGGG ATGGGGGATT ATCGAAATGA AACCTTGACG ACTAACCCGC

501 GCGACACTGC CTTTCTTTCC CACTTGGTAC AGACCGTATT TTTCCTGCGC  
 551 GGCATAGACG TTGTTTCTCC TGCCAATGCC GATACAGATG TGTTTATTAA  
 601 CATCGACGTA TTCGGAACGA TACGCAACAG AACCGAAATG..

This corresponds to the amino acid sequence <SEQ ID 78; ORF15>:

5 1 MQARLLIPIL FSVFILSACG TLTGIPSHGG XKRFAVEQEL VAASARAAVK  
 51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDAXXXG EYINSPAVRT  
 101 DYTYPRYETT AETTSGLTGG LTSLSTLNA PALSRQSDG SGSKSSLGLN  
 151 IGGMGDYRNE TLTTNPRDTA FLHLVQTVF FLRGIDVVSP ANADTDVFIN  
 201 IDVFGTIRNR TEM..

10 Further work revealed the complete nucleotide sequence <SEQ ID 79>:

1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC  
 51 CGCCTGCGGG AACTGACAG GTATTCCATC GCATGGCGGA GGTAAACGCT  
 101 TTGCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTAAAT  
 151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC  
 15 201 CACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA  
 251 TTGATGCACT GATTTCGTGGC GAATACATAA ACAGCCCTGC CGTCCGTACC  
 301 GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG  
 351 TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT  
 401 CTCGCACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT  
 451 ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGCG  
 20 501 CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATTT TTCCTGCGCG  
 551 GCATAGACGT TGTTCCTCCT GCCAATGCCG ATACAGATGT GTTTATTAAC  
 601 ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA  
 651 TGCCGAAACA CTGAAAGCCC AAACAAACT GGAATATTTC GCAGTAGACA  
 25 701 GAACCAATAA AAAATTGCTC ATCAAACCAA AAACCAATGC GTTTGAAGCT  
 751 GCCTATAAAG AAAATTACGC ATTGTGGATG GGGCCGTATA AAGTAAGCAA  
 801 AGGAATTAAA CCGACGGAAG GATTAAATGGT CGATTTCCTC GATATCCGAC  
 851 CATACGGCAA TCATACGGGT AACTCCGCCC CATCCGTAGA GGCTGATAAC  
 901 AGTCATGAGG GGTATGGATA CAGCGATGAA GTAGTGCAC AACATAGACA  
 30 951 AGGACAACCT TGA

This corresponds to the amino acid sequence <SEQ ID 80; ORF15-1>:

1 MQARLLIPIL FSVFILSACG TLTGIPSHGG GKRFAVEQEL VAASARAAVK  
 51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDALIRG EYINSPAVRT  
 101 DYTYPRYETT AETTSGLTGG LTSLSTLNA PALSRQSDG SGSKSSLGLN  
 151 IGGMGDYRNE TLTTNPRDTA FLHLVQTVF FLRGIDVVSP ANADTDVFIN  
 201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRNKKLL IKPKTNAFEA  
 251 AYKENYALWM GPYKVSKEIK PTEGLMVDFS DIRPYGNHTG NSAPSVEADN  
 301 SHEGYGYSDE VVRQHRQGP \*

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 81>:

40 1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC  
 51 CGCCTGCGGG AACTGACAG GTATTCCATC GCATGGCGGA GGTAAACGCT  
 101 TTGCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTAAAT  
 151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC  
 201 AACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA  
 45 251 TTGATGCACT GATTTCGTGGC GAATACATAA ACAGCCCTGC CGTCCGTACC  
 301 GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG  
 351 TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT  
 401 CGCGCACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT  
 451 ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGCG  
 50 501 CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATTT TTCCTGCGCG  
 551 GCATAGACGT TGTTCCTCCT GCCAATGCCG ATACGGATGT GTTTATTAAC  
 601 ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA  
 651 TGCCGAAACA CTGAAAGCCC AAACAAACT GGAATATTTC GCAGTAGACA  
 701 GAACCAATAA AAAATTGCTC ATCAAACCAA AAACCAATGC GTTTGAAGCT  
 55 751 GCCTATAAAG AAAATTACGC ATTGTGGATG GGACCGTATA AAGTAAGCAA  
 801 AGGAATTAAA CCGACGGAAG GATTAAATGGT CGATTTCCTC GATATCCAAC  
 851 CATACGGCAA TCATATGGGT AACTCTGCCC CATCCGTAGA GGCTGATAAC  
 901 AGTCATGAGG GGTATGGATA CAGCGATGAA GCAGTGCAC GACATAGACA  
 951 AGGGCAACCT TGA

60 This encodes a protein having amino acid sequence <SEQ ID 82; ORF15a>:

1 MQARLLIPIL FSVFILSACG TLTGIPSHGG GKRFAVEQEL VAASARAAVK

51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDALIRG EYINSPAVRT  
 101 DYTYPRYETT AETTSGGLTG LTSLSTLNA PALSRTQSDG SGSKSSLGLN  
 151 IGGMGDYRNE TLTNPRDTA FLSHLVQTVF FLRGIDVSP ANADTDVFIN  
 201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA  
 251 AYKENYALWM GPYKVSIGIK PTEGLMVDFF DIQPYGNHMG NSAPSVEADN  
 301 SHEGYGYSDE AVRRHRQGGP \*

The originally-identified partial strain B sequence (ORF15) shows 98.1% identity over a 213aa overlap with ORF15a:

10	orf15.pep	10 20 30 40 50 60	MQARLLIPILFSVFILSACGTLTGIPSHGGXKRFAVEQELVAASARA AVKMDLQALHGR
	orf15a	10 20 30 40 50 60	MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARA AVKMDLQALHGR
15	orf15.pep	70 80 90 100 110 120	KVALYIATMGDQSGSLTGGGRYSIDAXXGEYINSPAVRTDYTYPRYETTAETTSGGLTG
	orf15a	70 80 90 100 110 120	KVALYIATMGDQSGSLTGGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGGLTG
20	orf15.pep	130 140 150 160 170 180	LTSLSTLNAPALSRTQSDGSGSKSSLGLNIGGMGDYRNETLTNPRDTAFLSHLVQTVF
	orf15a	130 140 150 160 170 180	LTSLSTLNAPALSRTQSDGSGSKSSLGLNIGGMGDYRNETLTNPRDTAFLSHLVQTVF
25	orf15.pep	190 200 210	FLRGIDVSPANADTDVFINIDVFGTIRNRTEM
	orf15a	190 200 210 220 230 240	FLRGIDVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL

The complete strain B sequence (ORF15-1) and ORF15a show 98.8% identity in 320 aa overlap:

35	orf15a.pep	10 20 30 40 50 60	MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARA AVKMDLQALHGR
	orf15-1	10 20 30 40 50 60	MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARA AVKMDLQALHGR
40	orf15a.pep	70 80 90 100 110 120	KVALYIATMGDQSGSLTGGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGGLTG
	orf15-1	70 80 90 100 110 120	KVALYIATMGDQSGSLTGGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGGLTG
45	orf15a.pep	130 140 150 160 170 180	LTSLSTLNAPALSRTQSDGSGSKSSLGLNIGGMGDYRNETLTNPRDTAFLSHLVQTVF
	orf15-1	130 140 150 160 170 180	LTSLSTLNAPALSRTQSDGSGSKSSLGLNIGGMGDYRNETLTNPRDTAFLSHLVQTVF
50	orf15a.pep	190 200 210 220 230 240	FLRGIDVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
	orf15-1	190 200 210 220 230 240	FLRGIDVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
55	orf15a.pep	250 260 270 280 290 300	IKPKTNAFEAAAYKENYALWMGPYKVSIGIKPTEGLMVDFFDIQPYGNHMGNSAPSVEADN
	orf15-1	250 260 270 280 290 300	IKPKTNAFEAAAYKENYALWMGPYKVSIGIKPTEGLMVDFFDIRPYGNHTGNSAPSVEADN
60	orf15a.pep	310 320	SHEGYGYSDEAVRRHRQGGPX
	orf15-1	310 320	SHEGYGYSDEVVRQHRQGGPX

310

320

Further work identified the corresponding gene in *N.gonorrhoeae* <SEQ ID 83>:

```

1  ATGCGGGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
51 CGCCTGCGGG ACGCTGACAG GTATTCCATC GCATGGCGGA GGCAAACGCT
101 TCGCGGTGCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTTAAA
151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
201 AACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA
251 TTGATGCACT GATTGCGGCG GAATACATAA ACAGCCCTGC CGTCCGCACC
301 GATTACACCT ATCCGCGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
351 TTTGACGGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
401 CGCGCACCCA ATCAGACGGT AGCGGAAGTA GGAGCAGTCT GGGCTTAAAT
451 ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CCAACCCGCG
501 CGACACTGCC TTTCTTTCCC ACTTGGTGCA GACCGTATTT TTCCTGCGCG
551 GCATAGACGT TGTTCCTCCT GCCAATGCCG ATACAGATGT GTTTATTAAC
601 ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
651 TGCCGAAACA CTGAAAGCCC AAACAAACT GGAATATTTC GCAGTAGACA
701 GAACCAATAA AAAATGCTC ATCAAACCCA AAACCAATGC GTTTGAAGCT
751 GCCTATAAAG AAAATTACGC ATTGTGGATG GGGCCGTATA AAGTAAGCAA
801 AGGAATCAAA CCGACGGAAG GATTGATGGT CGATTTCTCC GATATCCAAC
851 CATACGGCAA TCATACGGGT AACTCCGCCC CATCCGTAGA GGCTGATAAC
901 AGTCATGAGG GGTATGGATA CAGCGATGAA GCAGTGCGAC AACATAGACA
951 AGGGCAACCT TGA

```

This encodes a protein having amino acid sequence <SEQ ID 84; ORF15ng>:

```

1  MRARLLIPIL FSVFILSACG TLTGIPSHGG GKRFVEQEL VAASARAANK
25 51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDALIRG EYINSPAVRT
101 DYTYPRYETT AETTSGLTGT LTSLSTLNA PALSRQSDG SGRSSLGLN
151 IGGMGDYRNE TLTNPRDTA FLSHLVQTVF FLRGIDVVP ANADTDVFIN
201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA
251 AYKENYALWM GPKVSKGIK PTEGLMVDFS DIQPYGNHTG NSAPSVEADN
301 SHEGYGSDE AVRQHRQGP *

```

The originally-identified partial strain B sequence (ORF15) shows 97.2% identity over a 213aa overlap with ORF15ng:

```

35 orf15.pep  MQARLLIPILFSVFILSACGTLTGIPSHGGKRFVEQELVAASARAANKDMDLQALHGR 60
   orf15ng   MRARLLIPILFSVFILSACGTLTGIPSHGGKRFVEQELVAASARAANKDMDLQALHGR 60
40 orf15.pep  KVALYIATMGDQSGSLTGGRYSIDAXXXGEYINSPAVRTDYTYPRYETTAETTSGLTGT 120
   orf15ng   KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGLTGT 120
45 orf15.pep  LTSLSTLNAPALSRQSDGSGSKSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF 180
   orf15ng   LTSLSTLNAPALSRQSDGSGSRSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF 180
50 orf15.pep  FLRGIDVVPANADTDVFINIDVFGTIRNRTEM 213
   orf15ng   FLRGIDVVPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL 240

```

The complete strain B sequence (ORF15-1) and ORF15ng show 98.8% identity in 320 aa overlap:

```

50 orf15-1.pep  10 20 30 40 50 60
   orf15ng     10 20 30 40 50 60
55 orf15-1.pep  70 80 90 100 110 120
   orf15ng     70 80 90 100 110 120
60 orf15-1.pep  130 140 150 160 170 180

```

indicates a putative leader sequence, and it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF15-1 (31.7kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 4A shows the results of affinity purification of the GST-fusion protein, and Figure 4B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 4C) and ELISA (positive result). These experiments confirm that ORFX-1 is a surface-exposed protein, and that it is a useful immunogen.

### Example 11

The following partial DNA sequence was identified in *N.meningitidis* <SEO ID 85>:

35	1	..GG.CAGCACA	AAAAACAGGC	GGTTGAACGG	AAAAACGTA	TTTACGATGA
	51	TGCCGGGTAT	GATATTCGGC	GTATTCACGG	GCGCATTCTC	CGCAAAATAT
	101	ATCCCCGCGT	TCGGGCTTCA	AATTTCTTC	ATCCTGTTTT	TAACCGCCGT
	151	CGCATTCAAA	ACACTGCATA	CCGACCCTCA	GACGGCATCC	CGCCCGCTGC
	201	CCGGACTGCG	CrGACTGACT	GCGGTTTCCA	CACGTGTCGG	CACAATGTCC
40	251	AGCTGGGTGC	GCATAGGCGG	CGGTTCACTT	TCCGTCCTCT	TCTTAATCCA
	301	CTGCGGCTTC	CCCGCCCAT	AAGCCATCGG	CACATCATCC	GGCCTTGCC
	351	GGCCGATTGC	ACTCTCCGGC	GCAATATCGT	ATCTGCTCAA	CGGCCTGAAT
	401	ATTGCAGGAT	TGCCCGAAGG	GTCATGGGC	TTCTTTTACC	TGCCCGCCGT
	451	CGCCGTCTCT	AGCGCGGCAA	CCATTGCCTT	TGCCCCGCTC	GGTGTCAAAA
45	501	CGGCCACAA	ACTTCTTCT	GCCAAACTCA	AAAAATC.TT	CGGCATTATG
	551	TTGCTTTTGA	TTGCGGGAAA	AATGCTGTAC	AACCTGCTTT	AA

This corresponds to the amino acid sequence <SEQ ID 86; ORF17>:

1 ..GQHKKQAVNG KTVFTMMPGM IFGVFTGAFS AKYIPAFGLQ IFFILELTAV  
51 AFKTLHTDPO TASRPLPLP XLTAVSTLEG TMSSWVGIGG GSLSVPFLIH

101 CGFPAHKAIG TSSGLAWPIA LSGAISYLLN GLNIAGLPEG SLGFLYLPVAV  
151 AVLSAATIAF APLGVKTAHK LSSAKLKKSF GIMLLLIAGK MLYNLL\*

Further work revealed the complete nucleotide sequence <SEQ ID 87>:

```

1  ATGTGGCATT GGGACATTAT CTTAATCCTG CTTGCCGTAG GCAGTGCGGC
5  51  AGGTTTTATT GCCGGCCTGT TCGGCGTAGG CGGCGGCACG CTGATTGTCC
101 CTGTCGTTTT ATGGGTGCTT GATTTCAGG GTTTGGCACA ACATCCTTAC
151 GCGCAACACC TCGCCGTCGG CACATCCTTC GCCGTCATGG TCTTCACCGC
201 CTTTTCCAGT ATGCTGGGGC AGCACAAAAA ACAGGCGGTC GACTGGAATA
251 CCGTATTAC  GATGATGCCG GGTATGATAT TCGGCGTATT CACGGGCGCA
10  301 CTCTCCGCAA AATATATCCC CGCGTTCGGG CTTCAAATTT TCTTCATCCT
351 GTTTTTTAACC GCCGTGCGAT TCAAAACACT GCATACCGAC CCTCAGACGG
401 CATCCCGCCC GCTGCCCGGA CTGCCCGGAC TGAATGCGGT TTCCACACTG
451 TTCGGCACAA TGTCGAGCTG GGTCCGCATA GGCGGCGGTT CACTTCCGT
501 CCCCTTCTTA ATCCACTGCG GCTTCCCGC CCATAAAGCC ATCGGCACAT
15  551 CATCCGGCCT TGCGTGGCCG ATTGCACTCT CCGGCGCAAT ATCGTATCTG
601 CTCAACGGCC TGAATATTGC AGGATTGCCC GAAGGGTCAC TGGGCTTCTT
651 TTACCTGCCC GCCGTGCGCG TCCTCAGCGC GGCAACCATT GCCTTTGCCC
701 CGCTCGGTGT CAAAACCGCC CACAAACTTT CTTCTGCCAA ACTCAAAAAA
751 Tc.TTCGGCA TTATGTTGCT TTTGATTGCC GGAAAAATGC TGTACAACCT
20  801 GCTTTAA

```

This corresponds to the amino acid sequence <SEQ ID 88; ORF17-1>:

```

1  MWHWDIILIL LAVGSAAGFI AGLFVGGGT LIVPVVLWVL DLQGLAQHPY
51  AQHLAVGTSF AVMVFAESS MLGQHKQAV DWKTVFTMMP GMIFGVFTGA
25  101 LSAKYIPAFG LQIFFILFLT AVAFKTLHTD PQASRPLPG LPGLTAVSTL
151 FGTMSWVG I GGGSLVPFL IHCGFPAHKA IGTSSGLAWP IALSGAISYL
201 LNGLNIAGLP EGS LGFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKK
251 XFGIMLLLIA GKMLYNLL*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical *H.influenzae* transmembrane protein HI0902 (accession number P44070)

30 ORF17 and HI0902 proteins show 28% aa identity in 192 aa overlap:

```

ORF17  3  HKKQAVNGKTVFTMMPGMIFGVFT-GAFSAKYIPAFGLQIF--FILFLTAVAFKTLHTDP 59
      HK  +  + V  + P ++ VF G F +      +IF  +++L      ++ D
HI0902 72  HKLGNIVQAVRILAPVIMLSVFICGLFIGRLDREISAKIFACLVVYLATKMVLSIKKD- 130

ORF17  60  QTASRPLPGLPXLTA VSTLFGTMSWVGIGGGSLVPFLIHCGFPAHKAIGTSSGLAWPI 119
      Q  ++ L L +      L G SS GIGG  VPFL  G  +AIG+S+  +
HI0902 131 QVTTKSLTPLSSVIG-GILIGMASSAAGIGGGFIVPFLTARGINIKQAIGSSAFCGMLL 189

ORF17  120 ALSGAISYLLNGLNIAGLPEGSLGFLYLPVAVLSAATIAFAPLVGXXXXXXXXXXXXXXXXX 179
      +SG S+++G      +PE SLG++YLPVAV ++A +      + LG
HI0902 190 GISGMFSFIVSGWGNPLMPEYSLGYIYLPVAVLGITATSFFTSKLGASATAKLPVSTLKKG 249

ORF17  180 FGIMLLLIAGKM 191
      F + L+++A M
HI0902 250 FALFLIVVAINM 261

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF17 shows 96.9% identity over a 196aa overlap with an ORF (ORF17a) from strain A of *N.*

*meningitidis*:

```

50  orf17.pep          10      20      30
      GQHKQAVNGKTVFTMMPGMIFGVFTGAFS
      |||||: |||||:||||:|
orf17a  QGLAQHPYAQHLAVGTSFAVMVFAESSMLGQHKQAVDWKTVFTMMPGMVFGVFAGALS
      50      60      70      80      90     100

55  orf17.pep          40      50      60      70      80      90
      AKYIPAFGLQIFFILFLTAVAFKTLHTDPQTASRPLPGLPXLTA VSTLFGTMSWVGIGG
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf17a  AKYIPAFGLQIFFILFLTAVAFKTLHTDPQTASRPLPGLPGLTAVSTLFGTMSWVGIGG

```

-103-

		110	120	130	140	150	160
		100	110	120	130	140	150
5	orf17.pep	GSLSVFPLIHCGFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPVAV					
	orf17a	GSLSVFPLIHCGFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPVAV					
		170	180	190	200	210	220
10	orf17.pep	160	170	180	190		
	orf17a	160	170	180	190		
		230	240	250	260		

The complete length ORF17a nucleotide sequence <SEQ ID 89> is:

15	1	ATGTGGCATT	GGGACATTAT	CTTAATCCTG	CTTGCCGTAG	GCAGTGGCGC
	51	AGGTTTTATT	GCCGGCCTGT	TCGGCGTAGG	CGGCGGCACG	CTGATTGTCC
	101	CTGTCGTTTT	ATGGGTGCTT	GATTTGCAGG	GTTTGGCACA	ACATCCTTAC
	151	GCGCAACACC	TCGCCGTCGG	CACATCCTTC	GCCGTCATGG	TCTTCACCGC
20	201	CTTTTCCAGT	ATGCTGGGGC	AGCACAAAAA	ACAGGCGGTC	GACTGGAAAA
	251	CCGTATTTAC	GATGATGCGG	GGTATGGTAT	TCGGCGTATT	CGCTGGCGCA
	301	CTCTCCGCAA	AATATATCCC	AGCGTTCGGG	CTTCAAATTT	TCTTCATCCT
	351	GTTTTTAACC	GCCGTCCGAT	TCAAAACACT	GCATACCGAC	CCTCAGACGG
	401	CATCCCGCCC	GCTGCCCGGA	CTGCCCGGAC	TGACTGCGGT	TTCCACACTG
25	451	TTCGGCACAA	TGTCGAGCTG	GGTCGGCATA	GGCGGCGGTT	CACTTTCGGT
	501	CCCTTCTTTA	ATCCACTGCG	GCTTCCCCGC	CCATAAAGCC	ATCGGCACAT
	551	CATCCGGCCT	TGCCTGGCCG	ATTGCACTCT	CCGGCGCAAT	ATCGTATCTG
	601	CTCAACGGCC	TGAATATTGC	AGGATTGCCG	GAAGGGTCAC	TGGGCTTCCT
	651	TTACCTGCCC	GCCGTCGCCG	TCCTCAGCGC	GGCAACCATT	GCCTTGGCCC
30	701	CGCTCGGTGT	CAAAACCGCC	CACAACTTT	CTTCTGCCAA	ACTCAAAAAA
	751	TCCTTCGGCA	TTATGTTGCT	TTTGATTGCC	GGAAAAATGC	TGTACAACCT
	801	GCTTTAA				

This encodes a protein having amino acid sequence <SEQ ID 90>:

	1	MWHWDIILIL	LAVGSAAGFI	AGLEGVGGGT	LIVPVVLWVL	DLQGLAQHPY
	51	AQHLAVGTSF	AVMVFTAFSS	MLGQHKQAV	DWKTFTMMP	GMVFGVFAGA
35	101	LSAKYIPAFG	LQIFFILELT	AVAFKTLHTD	PQTASRPLPG	LPGLTAVSTL
	151	FGTMSSWVGI	GGGSLSVFPL	IHCGFPAHKA	IGTSSGLAWP	IALSGAISYL
	201	LNGLNIAGLP	EGSLGFLYLP	AVAVLSAATI	AFAPLGVKTA	HKLSSAKLKK
	251	SEGMILLLIA	GKMLYNLL*			

ORF17a and ORF17-1 show 98.9% identity in 268 aa overlap:

40	orf17a.pep	10	20	30	40	50	60
	orf17-1	10	20	30	40	50	60
45	orf17a.pep	70	80	90	100	110	120
	orf17-1	70	80	90	100	110	120
50	orf17a.pep	130	140	150	160	170	180
	orf17-1	130	140	150	160	170	180
55	orf17a.pep	190	200	210	220	230	240
	orf17-1	190	200	210	220	230	240
60	orf17a.pep	250	260	269			
	orf17-1	250	260	269			

ORF17 shows 93.9% identity over a 196aa overlap with a predicted ORF (ORF17.ng) from *N. gonorrhoeae*:

	orf17.pep		GQHKKQAVNGKTVFTTMMPGMIFGVFTGAFS	30
			:   : :     : :	
10	orf17ng	QGLAQHPYAQHLAVGTSFAVMVFTAFSSMLGQHKKQAVDWKTFIFAMMPGMIFGVFAGALS		102
	orf17.pep	AKYIPAFGLQIFFILFLTAVAFKTLHTDPQTASRPLPGLPXLTAVSTLFGMTSSWVGIGG		90
		:    : :     : :		
	orf17ng	AKYIPAFGLQIFFILFLTAVAFKTLHTGRQTASRPLPGLPGLTAVSTLFGMSSWVGIGG		162
15	orf17.pep	GSLSVPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPVAV		150
		:    : :     : :		
	orf17ng	GSLSVPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLVNGLNIAGLPEGSLGFLYLPVAV		202
20	orf17.pep	AVLSAATIAFAPLGVKTAHKLSSAKLKKSFGIMLLLIAGKMLYNLL	196	
		:    : :     : :		
	orf17ng	AVLSAATIAFAPLGVKTAHKLSSAKLKESFGIMLLLIAGKMLYNLL	268	

An ORF17ng nucleotide sequence <SEQ ID 91> is predicted to encode a protein having amino acid sequence <SEQ ID 92>:

25 1 MWHWDIILIL LAVGSAGFI AGLFGVGGGT LIVPVVLWL DLQGLAQHPY  
51 AQHLAVGTSF AVMVFTAFSS MLGQHKQAV DWKTIFAMMP GMIFGVFAGA  
101 LSAKYIPAFG LQIFFILELT AVAFKTLHTG RQTASRPLPG LPGLTAVSTL  
151 FGAMSSWVGI GGGSLSVPF L IHCGFPAHKA IGTSSGLAWP IALSGAISYL  
201 VGNLNIAGLP EGSLGFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKE  
30 251 SFGIMLLLIA GKMLYNLL\*

Further work revealed the complete gonococcal DNA sequence <SEQ ID 93>:

	1	ATGTGGCATT	GGGACATTAT	CTTAATCCTG	CTTGCcgtag	gcAGTGGCGC
	51	AGGTTTTATT	GCCGGCCTGT	Tcgggtgtagg	cggcgGTACG	CTGATTGTCC
35	101	CTGTCTGTTTT	ATGGGTGCTT	GATTTGCAGG	GTTTGGCACA	ACATCCTTAC
	151	GCGCAACACC	TCGCGCTCGG	CacaTccttc	gcCGTCATGG	TCTTCACCGC
	201	CTTTTCCAGT	ATGTTGGGGC	AGCACAAAAA	ACAGGCGGTC	GACTGGA AAA
	251	CCATATTTGC	GATGATGCCG	GGTATGATAT	TCGGCGTATT	CGCTGGCGCA
	301	CTCTCCGCAA	AATATATCCC	CGCGTTCGGG	CTTCAAATTT	TCTTCACTCT
	351	GTTTTTAAAC	GCCGTCGCAT	TCAAAACACT	GCATACCGGT	CGTCAGACGG
40	401	CATCCCGCCC	GCTGCCCGGG	CTGCCCGGAC	TGACTGCGGT	TTCCCACTG
	451	TTCGGCGCAA	TGTCGAGCTG	GGTCGGCATA	GGCGGCGGTT	CACTTTCCGT
	501	CCCCTTCTTA	ATCCACTGCG	GCTTCCCCGC	CCATAAAGCC	ATCGGCACAT
	551	CATCCGGCCT	TGCCTGGCCG	ATTGCACTCT	CCGGCGCAAT	ATCGTATCTG
	601	GTCACCGGTC	TGAATATTGC	AGGATTGCCC	GAAGGGTCGC	TGGGCTTCCT
45	651	TTACCTGCCC	GCCGTCGCCG	TCCTCAGCGC	GGCAACCATT	GCCTTTGCCC
	701	CGCTCGGTGT	CAAAACCGCC	CACAAACTTT	CTTCTGCCAA	ACTCAAAGAA
	751	TCTTTCGGCA	TTATGTTGCT	TTTGATTGCC	GAAAAATGC	TGTACAACCT
	801	GCTTTAA				

This corresponds to the amino acid sequence <SEQ ID 94; ORF17ng-1>:

50 1 MWHWDIILIL LAVGSAAGFI AGLFGVGGGT LIVPVVLWVL DLQGLAQHPY  
51 AQHLAVGTSE AVMVFTAFSS MLGQHKQAV DWKTIFAMMP GMI FGVFAGA  
101 LSAKYIPAFG LQIFILEFLT AVAFKTLHTG RQTASRPLPG LPGLTAVSTL  
151 FGAMSSWVG I GGGSLSVPF L IHC GFPAHKA IGTSSGLAWP IALSGAISYL  
201 VNLNIAGLP EBSLGLFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKE  
55 251 SFGIMLLLIA GKMLYNLL\*

ORF17ng-1 and ORF17-1 show 96.6% identity in 268 aa overlap:

orf17-1.pep MWHWDIILILLAVGSAAGFIAGLFGVGGGTILIVPVVLWVLDLQGLAQHPYAQHLAVGTSF



-105-

	orfl7ng-1	 MWHWDIILILLAVGSAAGFIAGLFGVGGTLLIVPVVLWVLDLQGLAQHPYAQHLAVGTSF 10 20 30 40 50 60
5	orfl7-1.pep	70 80 90 100 110 120 AVMVFTAFSSMLGQHKQAVDWKTVFTMMPGMIFGVFTGALSAYIPAFGLQIFFILFLT      :
10	orfl7ng-1	70 80 90 100 110 120 AVMVFTAFSSMLGQHKQAVDWKTFAMMPGMIFGVFAGALSAYIPAFGLQIFFILFLT      :
15	orfl7-1.pep	130 140 150 160 170 180 AVAFKTLHTDPQTASRPLPGLPGLTAVSTLFGTMSSWVGIGGSSLSVPFLIHC GFPAHKA      :
20	orfl7ng-1	130 140 150 160 170 180 AVAFKTLHTGRQTASRPLPGLPGLTAVSTLFGAMSSWVGIGGSSLSVPFLIHC GFPAHKA      :
25	orfl7-1.pep	190 200 210 220 230 240 IGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPAAVAVLSAATIAFAPLGVKTA      :
30	orfl7ng-1	190 200 210 220 230 240 IGTSSGLAWPIALSGAISYLVNGLNIAGLPEGSLGFLYLPAAVAVLSAATIAFAPLGVKTA      :
35	orfl7-1.pep	250 260 269 HKLSSAKLKXKFGIMLLLIAGKMLYNLLX      :
40	orfl7ng-1	250 260 HKLSSAKLKESFGIMLLLIAGKMLYNLLX      :

In addition, ORF17ng-1 shows significant homology with a hypothetical *H. influenzae* protein:

30	sp P44070 Y902_HAEIN HYPOTHETICAL PROTEIN HI0902 pir  G64015 hypothetical protein HI0902 - Haemophilus influenzae (strain Rd KW20) gi 1573922 (U32772) H. influenzae predicted coding region HI0902 [Haemophilus influenzae]Length = 264 Score = 74 (34.9 bits), Expect = 1.6e-23, Sum P(2) = 1.6e-23 Identities = 15/43 (34%), Positives = 23/43 (53%)
35	Query: 55 AVGTSEFAVMVFTAFSSMLGQHKQAVDWKTFAMMPGMIFGVF 97 A+GTSEFA +V T S HK + W+ + + P ++ VF Sbjct: 52 ALGTSFATIVITGIGSAQRHHKLGNIWQAVRILAPVIMLSVF 94
40	Score = 195 (91.9 bits), Expect = 1.6e-23, Sum P(2) = 1.6e-23 Identities = 44/114 (38%), Positives = 65/114 (57%)
45	Query: 150 LFGAMSSWVGIGGSSLSVPFLIHC GFPAHKAIGTSSGLAWPIALSGAISYLVNGLNIAGL 209 L G SS GIGGG VPFL G +AIG+S+ + +SG S++V+G + Sbjct: 148 LIGMASSAAGIGGGGFIVPFLTARGINIKQAIGSSAFCGMLLGISGMFSFIVSGWGNPLM 207
50	Query: 210 PEGSLGFLYLPAAVAVLSAATIAFAPLGVKTAHKLSSAKLKESFGIMLLLIAGKM 263 PE SLG++YLPVAV ++A + + LG KL + LK+ F + L+++A M Sbjct: 208 PEYSLGYIYLPVAVLGITATSFTSKLGASATAKLPVSTLKKGFALFLIVVAINM 261

This analysis, including the homology with the hypothetical *H. influenzae* transmembrane protein, suggests that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 12

55 The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 95>:

1	..GGAAACGGAT GGCAGGCAGA CCCC GAACAT CCGCTGCTCG GGCTTTTTTC
51	CGTCAGTAAT GTATCGATGA CGCTTGCTTT TGTCCGAATA TGTGCGTTGG
101	TGCATTATTG CTTTCGGGA ACGGTTCAAG TGTTGTGTT TGCGGCACTG
151	CTCAAACTTT ATGCGCTGAA GCCGTTTAT TGGTTCGTGT TGCAGTTTGT
201	GCTGATGGCG GTTGCCATATG TCCACCGCTG CCGTATAGAC CGGCAGCCGC
60	251 CGTCAACGTT CGGCGCTCG CAGCTGCGAC TCGGCGGGT GACGGCAGCG

5

```

1  ..GNGWQADPEH PLLGLFAVSN VSMTLAFVGI CALVHYCFSG TVQVFVFAAL
51 LKLYALPKVY WFLVQFVLMA VAYVHRCGID RQPPSTFGGS QLRLLGLTAA
101 LMQVSVLVLL LSEIGR*

```

10

15

20

25

1	<u>MILLHLDFLS</u>	<u>ALLYAAVFLF</u>	<u>LIFRAGMLQW</u>	<u>FWASIMLWLG</u>	<u>ISVLGAKIMP</u>
51	<u>GIWGMTRAAP</u>	<u>LFIPHFYLT</u>	<u>LSIFFFIGHW</u>	<u>NRKTDGNQWQ</u>	<u>ADPEHLLGL</u>
101	<u>FAVSNVSMTL</u>	<u>AFVGICALVH</u>	<u>YCFSGTVQVF</u>	<u>VFAALLKLYA</u>	<u>LKPVYWEVLQ</u>
151	<u>FVLMAVAYVH</u>	<u>RCGIDRQPPS</u>	<u>TFGGSQLRLG</u>	<u>GLTAALMQVS</u>	<u>VLVLLLSEIG</u>
201	R*				

30

35

40

45

**meningitidis:**

```

                                10          20          30
orf18.pep                      GNGWQADPEHPLLGLFAVSNVSMTLAFVGI
                                |||
orf18a      TRAAPLFIPHFYLTGSIFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGI
                60          70          80          90          100         110

                                40          50          60          70          80          90
orf18.pep      CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAVAYVHRCGIDRQPPSTFGGS
                |||
orf18a      CALVHYCFSTVQVFVFAALLKLYALKPVYWFVLQFVLMAVAYVHRCGIDRQPPSTFGGS
                120         130         140         150         160         170

                                100          110
orf18.pep      QLRLGGLTAALMQSVSLVLLSEIGRX
                |||
orf18a      QLRLGGLTAALMQSVSLVLLSEIGRX
                180         190         200

```

50

55

1	ATGATTTTGC	TGCATTTGGA	TTTTTTGTCT	GCCTTACTGT	ATGCGGCGGT
51	TTTTCTGTTT	CTGATATTCC	GCGCAGGAAT	GTTGCAATGG	TTTGGGCGCA
101	GTATTATGCT	GTGGCTGGGC	ATATCGGTTT	TGGGGGCAAA	GCTGATGCCA
151	GCGATATGGG	GAATGACCCG	CGCCGCGCCC	TTGTTCATCC	CCCATTTTTA
201	CGTGACTTTG	GCGAGCATAT	TTTTTTTCAT	CGGGCATTGG	AACCGGAAAA
251	CGGATGGAAA	CGGATGGCAG	GCAGACCCCG	AACATCCTCT	GCTCGGGCTG
301	TTTGCCGTC A	GTAATGTATC	GATGACCGTT	GCTTTTGTGC	GAATATGTGC
351	GTTGGTGTCAT	TATTGCTTTT	CGNAGACGGT	TCAAGTGTTT	GTTGTTGCGG
401	CAC TGC TCAA	TCTTTATGCG	CTGAAGCCGG	TTTATTGGTT	CGTGTGTCAG

-107-

451 TTTGTGCTGA TGGCGGTTGC CTATGTCCAC CGCTGCGGTA TAGACCGGCA  
 501 GCCGCCGTCA ACGTTCGGCG GNTCGCAGCT GCGACTCGGC GGGTTGACGG  
 551 CAGCGTTGAT GCAGNTCTCG GTACTGGTGC TGCTGCTTTC AGAAATTGGA  
 601 AGATAA

5 This encodes a protein having amino acid sequence <SEQ ID 100>:

1 MILLHLDFLS ALLYAAVFLF LIFRAGMLQW FWASIMLWLG ISVLGAKLMP  
 51 GIWGMTRAAP LFIPHFYLTG GSIFFFIHGW NRKTDGNGWQ ADPEHPLLGL  
 101 FAVSNVSM TLAFVGICALVH YCFSXTVQVF VFAALLKLYA LKPVYWFVLQ  
 151 FVLMAYAYVH RCGIDRQPPS TFGGSQLRLG GLTAALMQXS VLVLLSEIG  
 201 R\*

ORF18a and ORF18-1 show 99.0% identity in 201 aa overlap:

		10	20	30	40	50	60
15	orf18a.pep	MILLHLDFLS	ALLYAAVFLF	LIFRAGMLQW	FWASIMLWLG	ISVLGAKLMP	GIWGMTRAAP
	orf18-1	MILLHLDFLS	ALLYAAVFLF	LIFRAGMLQW	FWASIMLWLG	ISVLGAKLMP	GIWGMTRAAP
		10	20	30	40	50	60
20	orf18a.pep	LFIPHFYLTG	SIFFFIHGW	NRKTDGNGWQ	ADPEHPLLGL	FAVSNVSM TL	AFVGICALVH
	orf18-1	LFIPHFYLTG	SIFFFIHGW	NRKTDGNGWQ	ADPEHPLLGL	FAVSNVSM TL	AFVGICALVH
		70	80	90	100	110	120
25	orf18a.pep	YCFSXTVQVF	VFAALLKLYA	LKPVYWFVLQ	FVLMAYAYVH	RCGIDRQPPS	TFGGSQLRLG
	orf18-1	YCFSXTVQVF	VFAALLKLYA	LKPVYWFVLQ	FVLMAYAYVH	RCGIDRQPPS	TFGGSQLRLG
		130	140	150	160	170	180
30	orf18a.pep	GLTAALMQXS	VLVLLSEIGR				
	orf18-1	GLTAALMQXS	VLVLLSEIGR				
35		190	200				

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF18 shows 93.1% identity over a 116aa overlap with a predicted ORF (ORF18.ng) from *N.gonorrhoeae*:

40	orf18.pep		GNGWQADPEHPLLGLFAVSNVSM TLAFVGI	30
	orf18ng	TRAAPLFIPHFYLTG	SIFFFIHGWNRKTDGNGWQADPEHPLLGLFAVSNVSM TLAFVGI	115
45	orf18.pep	CALVHYCFS	GTQVVFVFAALLKLYAL	90
	orf18ng	CALVHYCFS	GTQVVFVFAALLKLYAL	175
	orf18.pep	QLRLGGLTAALMQXS	VLVLLSEIGR	116
	orf18ng	QLRLGVLAA	MLMQVAVTAMLLAEIGR	201

50 The complete length ORF18ng nucleotide sequence is <SEQ ID 101>:

1 ATGATTTTGC TGCATTGGA TTTTGTCT GCCTTACTGt aTGCGGcgg  
 51 tttTctgTTT CTGATATCC GCGCAGGAAT GTTGCAATGG TTTTGGGCGA  
 101 GTATTGCGTT GTGGCTCGGC ATCTCGGTTT TAGGGGTAAA GCTGATGCCG  
 151 GGGATGTGGG GAATGACCCG CGCCGCGCCT TTGTTCATCC CCCATTTTAA  
 201 CCTGACTTTG GGCAGCATAT TTTTTCAT CCGGTATTGG AACCGGAAAA  
 251 CAGATGGAAA CGGATGGCAG GCAGACCCG AACATCCGCT GCTCGGGCTT  
 301 TTTGCCGTCA GTAATGTATC GATGACGCTT GCTTTTGTGCG GAATATGTGC  
 351 GTTGTGTCAT TATTGCTTTT CGGGAACGGT TCAAGTGTG GTGTTGCGG  
 401 CATTGCTCAA ACTTTATGCG CTGAAGCCGG TTTATTGGTT CGTGTGTCAG  
 451 TTTGTATTGA TGGCGGttgC CTATGTCCAC CGCTGCGGTA TAGACCGGCA  
 501 GCCGCCGTCA ACGTTCGGCG GTTCGCAGCT GCGACTCGGC GTGTTGCGCG

551 CGATGTTGAT GCAGGTTGCG GTAACGGCGA TGCTGCTTGC CGAAATCGGC  
601 AGATGA

This encodes a protein having amino acid sequence <SEQ ID 102>:

5  
1 MILLHLDFLS ALLYAAVFLE LIFRAGMLQW FWASIALWLG ISVLGVKLMP  
51 GMWGMTRAAP LFIPHFYLTLS GSIFFFIGYW NRKTDGNGWQ ADPEHPLLGL  
101 FAVSNVSMTL AFGVICALVH YCFSGTVQVF VFAALLKLYA LKPVYWFVLQ  
151 FVLMAYAVVH RCGIDRQPPS TFSGSQLRLG VLAAMLQVA VTAMLLAEIG  
201 R\*

This ORF18ng protein sequence shows 94.0% identity in 201 aa overlap with ORF18-1:

10	orf18-1.pep	10 20 30 40 50 60	MILLHLDFLSALLYAAVFLELIFRAGMLQWFWASIMLWLGISVLGAKLMPGIWGMTRAAP      :
	orf18ng	10 20 30 40 50 60	MILLHLDFLSALLYAAVFLELIFRAGMLQWFWASIALWLGISVLGVKLMPGMWGMTRAAP
15	orf18-1.pep	70 80 90 100 110 120	LFIPHFYLTLSGSIFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGVICALVH      :
	orf18ng	70 80 90 100 110 120	LFIPHFYLTLSGSIFFFIGYWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGVICALVH
20	orf18-1.pep	130 140 150 160 170 180	YCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAVVHRCGIDRQPPSTFGGSQLRLG      :
25	orf18ng	130 140 150 160 170 180	YCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAVVHRCGIDRQPPSTFGGSQLRLG
30	orf18-1.pep	190 200	GLTAALMQVSVLVLLSEIGRX  :     :  ::
	orf18ng	190 200	VLAAMLQVAVTAMLLAEIGRX

35 Based on this analysis, including the presence of several putative transmembrane domains in the  
gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and  
their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 13

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 103>:

40  
1 ATGAAAACCC CACTCCTCAA GCCTCTGCTN ATTACCTCGC TTCCCGTTTT  
51 CGCCAGTGTT TTTACGCGCG CCTCCATCGT CTGGCAGCTA GGCGAACCCA  
101 AGCTCGCCAT GCCCTTCGTA CTCGGCATCA TCGCCGGCGG CCTTGTCGAT  
151 TTGGACAACC NCNTGACCGG ACGGCTNAAA AACATCATCA CCACCGTCGC  
201 CCTGTTACCC CTCTCCTCGC TCACGGCACA AAGCACCCTC GGCACAGGGC  
251 TGCCCTTCAT CCTCGCCATG ACCCTGATGA CTT.CG.CTT CACCATTTTA  
45 301 GGCGCGGCG ...

This corresponds to the amino acid sequence <SEQ ID 104; ORF19>:

1 MKTPLLKPLL ITSLPVFASV FTAASIVWQL GERKLAMPFV LGIIAGGLVD  
51 LDNXTGRK NIITTVLFT LSSLTAQSTL GTGLPFILAM TLMTXXFTIL  
101 GAX...

50 Further work revealed the complete nucleotide sequence <SEQ ID 105>:

1 ATGAAAACCC CACTCCTCAA GCCTCTGCTC ATTACCTCGC TTCCCGTTTT  
51 CGCCAGTGTT TTTACGCGCG CCTCCATCGT CTGGCAGCTA GGCGAACCCA  
101 AGCTCGCCAT GCCCTTCGTA CTCGGCATCA TCGCCGGCGG CCTTGTCGAT  
151 TTGGACAACC GCCTGACCGG ACGGCTGAAA AACATCATCA CCACCGTCGC

201 CCTGTTTACC CTCTCCTCGC TCACGGCACA AAGCACCCCTC GGCACAGGGC  
 251 TGCCCTTCAT CCTCGCCATG ACCCTGATGA CCTTCGGCTT CACCATTTTA  
 301 GGCGCGGTG GGTCAAATA CCGCACCTTC GCCTTCGGTG CACTCGCCGT  
 351 CGCCACCTAC ACCACACTTA CCTACACCCC CGAAACCTAC TGGCTGACCA  
 401 ACCCCTTCAT GATTTTATGC GGCACCGTAC TGTACAGCAC CGCCATCCTC  
 451 CTGTTCCAAA TCGTCTGCC CCACCGCCCC GTCCAGAAA GCGTCGCCAA  
 501 CGCCTACGAC GCACTCGGCG GCTACCTCGA AGCCAAAGCC GACTTCTTCG  
 551 ACCCCGATGA GGCAGCCTGG ATAGGCAACC GCCACATCGA CCTCGCCATG  
 601 AGCAACACCG GCGTCATCAC CGCCTTCAAC CAATGCCGTT CCGCCCTGTT  
 651 TTACCGCCTT CGCGGCAAAC ACCGCCACCC GCGCACCGCC AAAATGCTGC  
 701 GTTACTACTT TGCCGCCCAA GACATACACG AACGCATCAG CTCCGCCAC  
 751 GTCGATTATC AGGAAATGTC CGAAAAATTC AAAAAACCG ACATCATCTT  
 801 CCGCATCCAC CGCCTGCTCG AAATGCAGGG ACAAGCCTGC CGCAACACCG  
 851 CCAAGCCCT GCGCGCAAGC AAAGACTACG TTTACAGCAA ACGCTCGGC  
 901 CGCGCCATCG AAGGTGCGG CCAATCGCTG CGCCTCCTTT CAGACAGCAA  
 951 CGACAGTCCC GACATCCGCC ACCTGCGCCG CTTTCTCGAC AACCTCGGCA  
 1001 CGGTGACCA GCAGTCCGC CAACTCCAGC ACAACGGCCT GCAGGCAGAA  
 1051 AACGACCGCA TGGGCGACAC CCGCATCGCC GCCCTCGAAA CCAGCAGCCT  
 1101 CAAAAACACC TGGCAGGCAA TCCGTCCGCA GCTAAACCTC GAATCAGGCG  
 1151 TATTCGCCA TGCCGTCCGC CTGTCCCTCG TCGTTGCCGC CGCCTGCACC  
 1201 ATCGTCGAAG CCCTCAACCT CAACCTCGGC TACTGGATAC TACTGACCGC  
 1251 CCTTTTCGTC TGCCAAACCA ACTACACCGC CACCAAAAGC CGCGTCCGCC  
 1301 AGCGCATCGC CGGCACCGTA CTCGGCGTAA TCGTCGGCTC GCTCGTCCCC  
 1351 TACTTCACCC CGTCTGTGCA AACCAAATC TGGATTGTCA TCGCCAGTAC  
 1401 CACCTCTTTT TTCATACCC GCACCTACAA ATACAGTTT TCCACCTTCT  
 1451 TCATTACCAT TCAAGCCCTG ACCAGCCTCT CCCTCGCAGG TTTGGACGTA  
 1501 TAGCCGCCA TGCCCGTACG CATCATCGAC ACCATTATCG GCGCATCCCT  
 1551 TGCTGGGCG GCAGTCAGCT ACCTGTGGCC AGACTGGAAA TACCTCACGC  
 1601 TCGAACGCAC CGCCGCCCTT GCCGTATGCA GCAACGGTGC CTATCTCGAA  
 1651 AAAATCACCG AACGCCCTCAA AAGCGGCGAA ACCGGGCGAC ACCTCGAATA  
 1701 CCGCGCCACC CGCCGCCGCG CCCACGAACA CACCGCGGCC CTCAGCAGCA  
 1751 CCCTTTCCGA CATGAGCAGC GAACCCGCAA AATTGCGCGA CAGCCTGCAA  
 1801 CCCGGCTTTA CCCTGTCAA AACCGGCTAC GCCCTGACCG GCTACATCTC  
 1851 CGCCCTCGGC GCATACCGCA GCGAAATGCA CGAAGATGC AGCCCGACT  
 1901 TTACCGCACA GTTCCACCTC GCGCGCAAC ACACCGCCCA CATCTTCCAA  
 1951 CACTGCCCC AAACCGAACC CGACGACTTT CAGACAGCAC TGGATACACT  
 2001 GCGCGCGGAA CTCGACACCC TCCGCACCCA CAGCAGCGGA ACACAAAGCC  
 2051 ACATCCTCCT CCAACAGCTC CAACTCATCG CCCGACAGT CGAACCTAC  
 2101 TACCGCGCCT ACCGCCAAAT TCCGCACAGG CAGCCCCAAA ATGCAGCCTG  
 2151 A

This corresponds to the amino acid sequence <SEQ ID 106; ORF19-1>:

1 MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD  
 51 LDNRLTGRK NIITVALEF LSSLTAQSTL GTGLPFILAM TLMTFGFTIL  
 101 GAVGLKYRTF AFGALAVATY TILTYTPETY WLTNPFMILC GTVLYSTAIL  
 151 LFQIVLPHRP VQESVANAYD ALGGYLEAKA DFFDPDEAAW IGNRHIDLAM  
 201 SNTGVITAFN QCRSALFYRL RGKRRHPRTA KMLRYFFAAQ DIHERISSAH  
 251 VDYQEMSEKF KNTDIIFRIH RLEMQGQAC RNTAQALRAS KDYVYSKRLG  
 301 RAIEGCRQSL RLLSDSNDSP DIRHLRRLD NLGSVDQFR QLQHNGLQAE  
 351 NDRMGDTRIA ALETSSLKNT WQAIRPQLNL ESGVFRHAVR LSLVVAACCT  
 401 IVEALNINLG YWILLTALFV CQPNYTATKS RVRQRIAGTV LGVIVGSLVP  
 451 YFTPSVETKL WIVIASTTLF FMTRTYKYSF STFFITIQL TSLSLAGLDV  
 501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLETAAL AVCSNGAYLE  
 551 KITERLKSGE TGDDVEYRAT RRAHEHTAA LSSTLSDMSS EPAKFADSLQ  
 601 PGFTLLKTY ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAEHTAHIFQ  
 651 HLPETEPDDF QTALDTRLGE LDTLRTHSSG TQSHILLQQL QLIARQLEPY  
 701 YRAYRQIPHR QPQNAA\*

Computer analysis of this amino acid sequence gave the following results:

Homology with predicted transmembrane protein YHFK of *H. influenzae* (accession number P44289)

ORF19 and YHFK proteins show 45% aa identity in 97 aa overlap:

60 orf19 6 LKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNXXTGRLKNIITT 65  
 L +I+++PVF +V AA +W +MP +LGIIAGGLVDLDN TGRKLN+ T  
 YHFK 5 LNAKVISTIPVFI AVNIAAVGIWFFDISSQSMPLILGIIAGGLVDLDNRLTGRKLNVEFT 64

orf19 66 VALFTLSSLTAQSTLGTGLPFIAMTLMTXFTILGA 102  
 + F++SS Q +G +I+ MT++T FT++GA  
 YHFK 65 LIAFSISSFIVQLHIGKPIQYIVLMTVLTFFIFTMIGA 101

# 5 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF19 shows 92.2% identity over a 102aa overlap with an ORF (ORF19a) from strain A of *N.*

*meningitidis*:

		10	20	30	40	50	60
10	orf19.pep	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLPMPFVLGIIAGGLVDLDNXXTGRLK					
	orf19a	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLPMPFVLGIIAGGLVDLDNRLTGRLK					
		10	20	30	40	50	60
15	orf19.pep	NIITTVALFTLSSLTAQSTLGTGLPFIAMTLMTXFTILGAX					
	orf19a	NIIATVALFTLSSLVAQSTLGTGLPFIAMTLMTFGFTIMGAVGLKYRTFAFGALAVATY					
		70	80	90	100	110	120
20	orf19a	TTLTYTPETYWLTNPFMILCGTVLYSTAILFQIILPHRPVQENVANAYEALGSYLEAKA					
		130	140	150	160	170	180

The complete length ORF19a nucleotide sequence <SEQ ID 107> is:

	1	ATGAAAACCC	CACCCCTCAA	GCCTCTGCTC	ATTACCTCGC	TTCCCGTTTT
	51	CGCCAGTGTC	TTTACCGCCG	CCTCCATCGT	CTGGCAGCTG	GGCGAACCCA
25	101	AGCTCGCCAT	GCCTTCGTA	CTCGGCATCA	TCGCTGGCGG	CCTGGTCGAT
	151	TTGGACAACC	GCCTGACCGG	ACGGCTGAAA	AACATCATCG	CCACCGTCGC
	201	CCTGTTCAAC	CTCTCCTCAC	TTGTGCGCGA	AAGCACCCCTC	GGCACAGGTT
	251	TGCCATTTCAT	CCTCGCCATG	ACCCTGATGA	CTTTCGGGCT	TACCATCATG
	301	GGCGCGGTCG	GGCTGAAATA	CCGCACCTTC	GCCTTCGGCG	CACTCGCCGT
30	351	CGCCACCTAC	ACCACACTTA	CCTACACCCC	CGAAACCTAC	TGGCTGACCA
	401	ACCCCTTTAT	GATTCGTGTC	GGAACCGTAC	TGTACAGCAC	CGCCATCATC
	451	CTGTTCCAAA	TCATCTGCC	CCACCGCCCC	GTTCAAGAAA	ACGTCGCCAA
	501	CGCCTACGAA	GACTCGGGCA	GCTACCTCGA	AGCCAAAGCC	GACTTTTTTCG
	551	ATCCCGACGA	AGCCGAATGG	ATAGGCAACC	GCCACATCGA	CCTCGCCATG
35	601	AGCAACACCG	GCGTCATCAC	CGCCTTCAAC	CAATGCCGTT	CCGCCCTGTT
	651	TTACCGCCTT	CGCGGCAAAC	ACCGCCACCC	GCGCACCGCC	AAAATGCTGC
	701	GCTACTACTT	CGCCGCCCAA	GACATACACG	AACGCATCAG	CTCCGCCAC
	751	GTCGACTACC	AAGAGATGTC	CGAAAAATTC	AAAAACACCG	ACATCATCTT
	801	CCGCATCCAC	CGCCTGCTCG	AAATGCAGGG	ACAAGCCTGC	CGCAACACCG
40	851	CCCAAGCCCT	GCGCGCAAGC	AAAGACTACG	TTTACAGCAA	ACGCCTCGGC
	901	CGCGCCATCG	AAGGCTGCCG	CCAATCGCTG	CGCCTCCTTT	CAGACAGCAA
	951	CGACAATCCC	GACATCCGCC	ACCTGCGCCG	CCTTCTCGAC	AACCTCGGCA
	1001	GCGTCGACCA	GCACTTCCGC	CAACTCCAGC	ACAACGGCCT	GCAGGCAGAA
	1051	AACGACCGCA	TGGCGGACAC	CCGCATCGCC	GCCCTCGAAA	CCGGCAGCCT
45	1101	CAAAAACACC	TGGCAGGCAA	TCCGTCCGCA	GCTAAACCTC	GAATCAGGCG
	1151	TATTCGCGCA	TGCCGTCCGC	CTGTCCCTTG	TCGTTGCCGC	CGCCTGCACC
	1201	ATCGTCGAAG	CCCTCAACCT	CAACCTCGGC	TACTGGATAC	TACTGACCGC
	1251	CCTTTTCGTC	TGCCAACCCA	ACTACACCGC	CACCAAAGC	CGCGTCCGCC
	1301	AGCGCATCGC	CGGCACCGTA	CTCGGCGTAA	TCGTCCGGTC	GCTCGTCCCC
50	1351	TACTTTACCC	CCTCCGTCGA	AACCAAATC	TGGATCGTCA	TCGCCAGTAC
	1401	CACCCCTCTT	TTCATGACCC	GCACCTACAA	ATACAGCTTC	TCGACATTTT
	1451	TCATCACCAT	TCAAGCCCTG	ACCAGCCTCT	CCCTCGCAGG	GTTGGACGTA
	1501	TACGCCGCCA	TGCCCGTACG	CATCATCGAC	ACCATTATCG	GCGCATCCCT
	1551	TGCCTGGGCG	GCACTCAGCT	ACCTGTGGCC	AGACTGGAAA	TACCTCACGC
55	1601	TCGAACGCAC	CGCCGCCCTT	GCCGTATGCA	GCAACGGCGC	CTATCTCGAA
	1651	AAAATCACCG	AACGCCTCAA	AAGCGGCGAA	ACCGGCGACG	ACGTCGAATA
	1701	CCGCGCCACC	CGCCGCCGCG	CCCACGAACA	CACCGCCGCC	CTCAGCAGCA
	1751	CCCTTTCCGA	CATGAGCAGC	GAACCGCAA	AATTCGCCGA	CAGCCTGCAA
	1801	CCCGGCTTTA	CCCTGCTCAA	AACCGGTAC	GCCCTGACCG	GCTACATCTC
60	1851	CGCCCTCGGC	GCATACCGCA	GCGAAATGCA	CGAAGAATGC	AGCCCCGACT
	1901	TTACCGCACA	GTTCCACCTC	GCCGCGGAAC	ACACCGCCCA	CATCTTCCAA
	1951	CACCTGCCCG	AAACCGAACC	CGACGACTTT	CAGACAGCAC	TGGATACACT
	2001	GCGCGGCGAA	CTCGACACCC	TCCGCACCCA	CAGCAGCGGA	ACACAAGGCC
	2051	ACATCCTCCT	CCAACAGCTC	CAACTCATCG	CCCGGCAGCT	CGAACCTTAC
65	2101	TACCGCGCCT	ACCGACAAAT	TCCGCACAGG	CAGCCCCAAA	ACGCAGCCTG
	2151	A				

This encodes a protein having amino acid sequence <SEQ ID 108>:

```

1  MKTPPLKPLL ITSLPVFASV FTAASIVWQL GEPKlampfv LGIIAGGLVD
51  LDNRLTGRlk NIIATVALFT LSSlVAQSTL GTGLPFILAM TlMTFGFTIM
101 GAVGLKYRTF AFGALAVATY TlTYTPETY WLTNPfMILC GtVlySTAIi
5  151  LFQIILPHRP VQENVANAYE ALGSYLEAKA DFFDPDEAEW IGnRHIDLAM
201  SNTGVITAFN QCRSALEFYL RGKhrHPRTA KMLRYyFAAQ DIHERISSAH
251  VDYQEMSEKF KNTDIIFRIH RLEMQGQAC RNTAQALRAS KDyVYSKRLG
301  RAIEGCRQSL RLLSDSNDNP DIRHLRRLD NLGSVDQQFR QlQHnGLQAE
10  351  NDRMGDTRIA ALETGSLKNT WQAIrPQLNL ESGVFRHAVR LSLVVAaACT
401  IVEALNLNLG YWILLTALFV CQPNyTATKS RVRQRiAGTV LGViVGSLVp
451  YFTPSVETKL WIViASTTLF FMTrTYKYSF STFFITiQAL TSLSLAGLDV
501  YAAMPVRIID TIIGASLAWA AVSYLWPdWK YLTlERTaAL AVCSNGAYLE
551  KITERLKSGE TGDDVEYRAT RRAHEHTAA LSSTLSdMSS EpaKFADSLQ
15  601  PGFTLLKTGY ALTGYISALG AYrSEMHEEC SPdFTAQFHL AAEHTAHIFQ
651  HLPETEPDDF QTALDtlRGE LDtLRTHSSG TQSHILLQQL QLIARQLEPY
701  YRAYRQIPHR QPQNAA*

```

ORF19a and ORF19-1 show 98.3% identity in 716 aa overlap:

```

20  orf19a.pep      10      20      30      40      50      60
    MKTPPLKPELLITSLPVFASVFTAASIVWQLGEPKlampfvLGIIAGGLVDLDNRLTGRlk
    orf19-1        10      20      30      40      50      60
    MKTPPLKPELLITSLPVFASVFTAASIVWQLGEPKlampfvLGIIAGGLVDLDNRLTGRlk

25  orf19a.pep      70      80      90      100     110     120
    NIIATVALFTLSSlVAQSTLGTGLPFILAMTlMTFGFTIMGAVGLKYRTFAFGALAVATY
    orf19-1        70      80      90      100     110     120
    NIITTVAlFTLSSlTAQSTLGTGLPFILAMTlMTFGFTILGAVGLKYRTFAFGALAVATY

30  orf19a.pep      130     140     150     160     170     180
    TTLTYTPETYWLTNPfMILCGTVlySTAIiLFQIILPHRPVQENVANAYEALGSYLEAKA
    orf19-1        130     140     150     160     170     180
    TTLTYTPETYWLTNPfMILCGTVlySTAIiLLFQIVLPHRPVQESVANAYDALGGYLEAKA

35  orf19a.pep      190     200     210     220     230     240
    DFFDPDEAEWIGNRHIDLAMsNTGVITAFNQCRSALEFYLRGKhrHPRTAKMLRYyFAAQ
    orf19-1        190     200     210     220     230     240
    DFFDPDEAAWIGNRHIDLAMsNTGVITAFNQCRSALEFYLRGKhrHPRTAKMLRYyFAAQ

40  orf19a.pep      250     260     270     280     290     300
    DIHERISSAHVDYQEMSEKFkNTDIIFRIHRLLEMQGQACRNTAQALRASKDYVYSKRLG
    orf19-1        250     260     270     280     290     300
    DIHERISSAHVDYQEMSEKFkNTDIIFRIHRLLEMQGQACRNTAQALRASKDYVYSKRLG

50  orf19a.pep      310     320     330     340     350     360
    RAIEGCRQSLRLLSDSNDNPDIRHLRRLDNLGSVDQQFRQLQHnGLQAEndRMGDTRIA
    orf19-1        310     320     330     340     350     360
    RAIEGCRQSLRLLSDSNDSPDIRHLRRLDNLGSVDQQFRQLQHnGLQAEndRMGDTRIA

55  orf19a.pep      370     380     390     400     410     420
    ALETGSLKNTWQAIrPQLNLESGVFRHAVRSLSVVAaACTIVEALNLNLGYWILLTALFV
    orf19-1        370     380     390     400     410     420
    ALETSSLKNTWQAIrPQLNLESGVFRHAVRSLSVVAaACTIVEALNLNLGYWILLTALFV

60  orf19a.pep      430     440     450     460     470     480
    CQPNyTATKSrVRQRiAGTVLGViVGSLVpYFTPSVETKLWIViASTTLFfMTrTYKYSF
    orf19-1        430     440     450     460     470     480
    CQPNyTATKSrVRQRiAGTVLGViVGSLVpYFTPSVETKLWIViASTTLFfMTrTYKYSF

65  orf19a.pep      490     500     510     520     530     540
    STFFITiQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSylWPdWKYLTlERTaAL

```

5	orf19-1	STFFITIQA	LTSLAGLD	VYAAMPV	RIIDTIIG	ASLAWAA	VSYLWPD	WKYLT	LERTAAL
		490	500	510	520	530	540		
10	orf19a.pep	550	560	570	580	590	600		
	orf19-1	AVCSNGAYLEKITERLKSGETGDDVEYRATRRRAHEHTAALSSTLS	SDMSSEPAKFADSLQ						
15	orf19a.pep	610	620	630	640	650	660		
	orf19-1	PGFTLLKGTGYALTGYISALGAYRSEMHEECSPDFTAQFH	LAAEHTAHIFQHL	PETEPDDF					
20	orf19a.pep	670	680	690	700	710			
	orf19-1	QTALDTLRGELDTLRTHSSGTQSHILLQQLIARQLEPYR	RAYRQIPHRQP	QNAAX					

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF19 shows 95.1% identity over a 102aa overlap with a predicted ORF (ORF19.ng) from *N.gonorrhoeae*:

25	orf19.pep	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNXXTGR	60
	orf19ng	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGR	60
30	orf19.pep	NIITTVAFTLSSSLTAQSTLGTGLPFILAMTLMTXXFTILGAX	103
	orf19ng	NIIATVAFTLSSSLTAQSTLGTGLPFILAMTLMTFGFTILGAVGLKYRTFAFGALAVATY	120

An ORF19ng nucleotide sequence <SEQ ID 109> is predicted to encode a protein having amino acid sequence <SEQ ID 110>:

35	1	MKTPLLKPLL	ITSLPVFASV	FTAASIVWQL	GEPKLAMPFV	LGIIAGGLVD
	51	LDNRLTGR	LK NIIATVAFT	LSSSLTAQSTL	GTGLPFILAM	TLMTFGFTIL
40	101	GAVGLKYRTF	AFGALAVATY	TTLTYTPETY	WLTNPFMILC	GTVLYSTAI
	151	LFQIILPHRP	VQESVANAYE	ALGGYLEAKA	DDFDPDEAAW	IGNRHIDLAM
45	201	SNTGVITAFN	QCRSALFYRL	RGKHRHPRTA	KMLRYYFAAQ	DIHERISSAH
	251	VDYQEMSEKF	KNTDIIFRIR	RLLEMQGOAC	RNTAQAIRSG	KDYVYSKRLG
50	301	RAIEGCRQSL	RLSDGNDSP	DIRHLSRLLD	NLGSDVQQFR	QLRHSDSPA
	351	NDRMGDTRIA	ALETGSFKNT	*		

Further work revealed the complete nucleotide sequence <SEQ ID 111>:

45	1	ATGAAACCC	CACTCCTCAA	GCCTCTGCTC	ATTACCTCGC	TTCCCGTTTT
	51	CGCCAGTGTC	TTTACCGCCG	CCTCCATCGT	CTGGCAGCTA	GGCGAACCCA
50	101	AGCTCGCCAT	GCCCTTCGTA	CTCGGCATCA	TCGCCGCGCG	CCTGGTCGAT
	151	TTGGACAACC	GCCTGACCGG	ACGGCTGAAA	AACATCATCG	CCACCGTCGC
55	201	CCTGTTTACC	CTCTCCTCGC	TCACGGCGCA	AAGCACCCTC	GGCACAGGGC
	251	TGCCCTTCAT	CCTCGCCATG	ACCCTGATGA	CCTTCGGCTT	TACCATTTTA
60	301	GGCGCGGTCG	GGCTGAAATA	CCGCACCTTC	GCCTTCGGCG	CACTCGCCGT
	351	CGCCACCTAC	ACCACGCTTA	CCTACACCCC	CGAAACCTAC	TGGCTGACCA
65	401	ACCCCTTCAT	GATTTTATGC	GGCACCGTAC	TGTACAGCAC	CGCCATCATC
	451	CTGTTCCAAA	TCATCCTGCC	CCACCGCCCC	GTCCAAGAAA	GCGTCGCCAA
70	501	TGCCTACGAA	GCACTCGGCG	GCTACCTCGA	AGCCAAAGCC	GACTTCTTCG
	551	ACCCGATGA	GGCAGCCTGG	ATAGGCAACC	GCCACATCGA	CCTCGCCATG
75	601	AGCAACACCG	GCGTCATCAC	CGCCTCAAC	CAATGCCGTT	CCGCCCTGTT
	651	TTACCGTTTG	CGCGGCAAA	ACCGCCACCC	GCGCACCGCC	AAAATGCTGC
80	701	GCTACTACTT	CGCCGCCCAA	GACATCCACG	AACGCATCAG	CTCCGCCAC
	751	GTCGACTACC	AAGAGATGTC	CGAAAAATTC	AAAAACACCG	ACATCATCTT
85	801	CCGCATCCGC	CGCCTGCTCG	AAATGCAGGG	GCAGGCGTGC	CGCAACACCG
	851	CCCAAGCCAT	CCGGTCGGGC	AAAGACTAcg	tTTACAGCAA	ACGCCTCGGA
90	901	CGCGCATCg	aaggctgCCG	CCAGTCGCTg	cgcctCCTTt	cagacggcAa
	951	CGACAGTCCC	GACATCCGCC	ACCTGAGccg	CCTTCTCGAC	AACCTCGgca



1001 GCGTcgacca gcagtTCcgc caactCCGAC ACAGcgactC CCCCCGcgaa  
 1051 Aacgaccgca tgggcgacaC CCGCATCGCC GCCCtcgaaa ccggcagctT  
 1101 caaaaaCAcc tggcaggCAA TCCGTCCGCa gctgaaCTC GAATCatgCG  
 1151 TATTCGCGCA TGCCGTCCGC CTGTCCCTCG TCGTTGCCGC CGCCTGCACC  
 5 1201 ATCGTCgaag cCCTCAACCT CAACCTCGGC TACTGGATAC TGCTGACCGC  
 1251 CCTTTTCGTC TGCCAACCCA ACTACACCGC CACCAAAGC CGCGTGATCC  
 1301 AACGCATCGC CGGCACGTA CTCGGCGTAA TCGTCGGCTC GCTCGTCCCC  
 1351 TACTTCACCC CCTCCGTCGA AACCAAACCT TGGATTGTCA TCGCCGGTAC  
 1401 CACCCTGTTC TTCATGACCC GCACCTACAA ATACAGTTTC TCCACCTTCT  
 10 1451 TCATCACCAT TCAGGCACTG ACCAGCCTCT CCCTCGCAGG TTTGGACGTA  
 1501 TACGCCGCCA TGCCCGTGCG CATCATcgac ACCATTATCG GCGCATCCCT  
 1551 TGCTGGGCG GCGGTCACT ACCTGTGGCC AGACTGGAAA TACCTCACGC  
 1601 TCGAACGCAC CGCCGCCCTT GCCGTATGCA GCAGCGGCAC ATACCTCCAA  
 1651 AAAATTGCCG AACGCCTCAA AACCGGCGAA ACCGGCGACG ACATAGAATA  
 15 1701 CCGCATCACC CGCCGCCGCG CCCACGAACA CACCGCGGCC CTCAGCAGCA  
 1751 CCCTTTCGGA CATGAGCAGC GAACCCGCAA AATTCGCGCA CAGCTGCAA  
 1801 CCGGCTTTA CCCTGTCAA AACCGGTAC GCCGTGACCG GCTACATCTC  
 1851 CGCCTCGGC GCATACCGCA GCGAAATGCA CGAAGAATGC AGCCCCGACT  
 1901 TTACCGCACA GTTCCACCTT GCCGCCGAAC ACACCGCCA CATCTTCAA  
 20 1951 CACCTGCCCG ACATGGGACC CGACGACTTT CAGACGGCAT TGGATACACT  
 2001 GCGCGGCGAA CTCGGCACCC TCCGCACCCG CAGCAGCGGA ACACAAAGCC  
 2051 ACATCCTCCT CCAACAGCTC CAACTCATCG CccgGCAACT CGAACCTTAC  
 2101 TACCGCGCCT ACCGACAAAT TCCGCACAGG CAGCCCCAAA ACGCAGCCTG  
 2151 A

25 This corresponds to the amino acid sequence <SEQ ID 112; ORF19ng-1>:

1 MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD  
 51 LDNRLTGRLL NIIATVALET LSSSLTAQSTL GTGLPFILAM TLMTEGFTIL  
 101 GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAIL  
 151 LFQIILPHRP VQESVANAYE ALGGYLEAKA DFFDPDEAAW IGNRHIDLAM  
 30 201 SNTGVITAFN QCRSALFYRL RGKHRHPRTA KMLRYFYAAQ DIHERISSAH  
 251 VDYQEMSEKF KNTDIIIFRI RLEMQGQAC RNTAQAIRSG KDYVYSKRLG  
 301 RATEGCRQSL RLLSDGNDSP DIRHLSRLLD NLGSVDQQFR QLRHSDSPAE  
 351 NDRMGDTRIA ALETGSFKNT WQAIRPQLNL ESCVFRHAVR LSLVVAACCT  
 401 IVEALNLNLG YWILLTALFV CQPNYTATKS RVYQRIAGTV LGVIVGSLVP  
 35 451 YFTPSVETKL WIVIAAGTTF FMTRTYKYSE STFFITIQAL TSLSLAGLDV  
 501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLETAAL AVCSSGTYLQ  
 551 KIAERLKTGE TGDDIEYRIT RRAHEHTAA LSSTLSDMSS EPAKFADSLQ  
 601 PGFTLLKTGY ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAEHTAHIFQ  
 651 HLPDMGPDDF QTALDTRLGE LGTLRTRSSG TQSHILLQQL QLIARQLEPY  
 40 701 YRAYRQIPHR QPQNAA\*

ORF19ng-1 and ORF19-1 show 95.5% identity in 716 aa overlap:

		10	20	30	40	50	60
45	orf19-1.pep	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLL					
	orf19ng-1	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLL					
		10	20	30	40	50	60
50	orf19-1.pep	NIIITVALEFTLSSSLTAQSTLGTGLPFILAMTLMTEGFTILGAVGLKYRTFAFGALAVATY					
	orf19ng-1	NIIATVALEFTLSSSLTAQSTLGTGLPFILAMTLMTEGFTILGAVGLKYRTFAFGALAVATY					
		70	80	90	100	110	120
55	orf19-1.pep	TTLTYTPETYWLTNPFMILCGTVLYSTAILLFQIIVLPHRPVQESVANAYDALGGYLEAKA					
	orf19ng-1	TTLTYTPETYWLTNPFMILCGTVLYSTAILLFQIILPHRPVQESVANAYEALGGYLEAKA					
		130	140	150	160	170	180
60	orf19-1.pep	DFFDPDEAAWIGNRHIDLAMSNITGVITAFNQCRSALFYRLRGKHRHPRTAKMLRYFYAAQ					
	orf19ng-1	DFFDPDEAAWIGNRHIDLAMSNITGVITAFNQCRSALFYRLRGKHRHPRTAKMLRYFYAAQ					
		190	200	210	220	230	240
65	orf19-1.pep	DIHERISSAHVDYQEMSEKFKNTDIIIFRIHRLLEMQGQACRNTAQALRASKDYVYSKRLG					
		250	260	270	280	290	300

orf19ng-1	DIHERISSAHVDYQEMSEKFKNTDIIFRIRRLLEMGGQACRNTAQAIRSGKDYVYSKRLG	250	260	270	280	290	300
orf19-1.pep	RAIEGCRQSLRLSDSNDSPDIRHLRRLLDNLGSVDQQFRQLQHNGLQAENDRMGDTRIA	310	320	330	340	350	360
orf19ng-1	RAIEGCRQSLRLSDGNDSPDIRHLSRLLDNLGSVDQQFRQLRHSDSPAENDRMGDTRIA	310	320	330	340	350	360
orf19-1.pep	ALETSSLKNTWQAIRPQLNLESGVFRHAVRLSLVVAAACTIVEALNLNLGYWILLTALFV	370	380	390	400	410	420
orf19ng-1	ALETGSFKNTWQAIRPQLNLESCVFRHAVRLSLVVAAACTIVEALNLNLGYWILLTALFV	370	380	390	400	410	420
orf19-1.pep	CQPNYTATKSRVRQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSF	430	440	450	460	470	480
orf19ng-1	CQPNYTATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSF	430	440	450	460	470	480
orf19-1.pep	STFFITIQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTTLRTAAL	490	500	510	520	530	540
orf19ng-1	STFFITIQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTTLRTAAL	490	500	510	520	530	540
orf19-1.pep	AVCSNGAYLEKITERLKSGETGDDVEYRATRRRAHEHTAALSSTLSDMSSEPAKFADSLQ	550	560	570	580	590	600
orf19ng-1	AVCSSGTYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSTLSDMSSEPAKFADSLQ	550	560	570	580	590	600
orf19-1.pep	PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHLAAEHTAHIFQHLPETEPDDEF	610	620	630	640	650	660
orf19ng-1	PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHLAAEHTAHIFQHLPMGPDDEF	610	620	630	640	650	660
orf19-1.pep	QTALDTRLRGELDTLRTHSSGTQSHILLQQLQLIARQLEPYRAYRQIPHROPQNAAX	670	680	690	700	710	
orf19ng-1	QTALDTRLRGELGTLRTRSSGTQSHILLQQLQLIARQLEPYRAYRQIPHROPQNAAX	670	680	690	700	710	

In addition, ORF19ng-1 shows significant homology to a hypothetical gonococcal protein previously entered in the databases:

sp|O33369|YOR2\_NEIGO HYPOTHETICAL 45.5 KD PROTEIN (ORF2) gnl|PID|e1154438  
(AJ002423) hypothetical protein [Neisseria gonorrh] Length = 417  
Score = 1512 (705.6 bits), Expect = 5.3e-203, P = 5.3e-203  
Identities = 301/326 (92%), Positives = 306/326 (93%)

Query: 307 RQSLRLLSDGNDSPDIRHLSRLLDNLGSVDQQFQRLRHSDSPAENDRMGDTRIAALETGS 366  
RQSLRLLSDGNDSDIRHLSRLLDNLGSVDQQFQRLRHSDSPAENDRMGDTRIAALETGS  
Sbjct: 1 RQSLRLLSDGNDSDIRHLSRLLDNLGSVDQQFQRLRHSDSPAENDRMGDTRIAALETGS 60

Query: 367 FKNTWQAIRPQLNLESCVFRHAVRLSLVVAAACTIVEALNINLGYWILLTALFVCQPNYT 426  
FKNTWQAIRPQLNLES VFRHAVRLSLVVAAACTIVEALNINLGYWILLT LFVCQPNYT  
Sbjct: 61 FKNTWQAIRPQLNLESGVFRHAVRLSLVVAAACTIVEALNINLGYWILLTRFVCQPNYT 120

Query: 427 ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIAGTTLFFMTRTRYKYSFSTFFIT 486  
ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIAGTTLFFMTRTRYKYSFSTFFIT  
Sbjct: 121 ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIAGTTLFFMTRTRYKYSFSTFFIT 180

Query: 487 IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTERTAAALAVCSSG 546  
 IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTERTAAALAVCSSG  
 Sbjct: 181 IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTERTAAALAVCSSG 240

Query: 547 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSSTLSDMSSEPAKFADSLQPGFTLL 606  
 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSSTLSDMSSEPAKFAD+ P  
 Sbjet: 241 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSSTLSDMSSEPAKFADTCNPALPCS 300

5 Query: 607 KTGyALTGYISALGAYRSEMHEECSP 632  
 K ALTGYISALG ++ + +P  
 Sbjet: 301 KPATALTYISALGHATAAKCTKNAAP 326

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein (the first of which is also seen in the meningococcal protein), and on homology with the YHFK protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 14

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 113>:

```

15      1 ATGAATATGC TGGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTC
      51 GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTGCGCG
    101 CGGGTATGGC GACGGATGCG TTTTGTGTCG CGTTCAAACCT GCCCAACCTG
      151 CTTCCGCGCG TGTTTGCGGA GGGGCGGTTT GCCCAAGCGT TTGTGCCGAT
    201 TTTGGCGGAA TACAAGGAAA CGCGTTCAAA AGAGGCGG.C GAAGCCTTTA
    251 TCCGCCATGT GCGGGGATG CTGTCGTTTG TACTGGTTAT CGTTACCGCG
    301 CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTTCCG CACCCGAGTT
    351 TTGCCCAAGA TGCCGACAAA TTTCAGCTCT CCATCGATTG GCTGCGGATT
    401 ACGTTTCCTT ATATATTATT GATTTCCTTG TCTTCATTG TCGGCTCGGT
    451 ACTCAATTCT TATCATAAGT TCGGCATTCC GCGGTTTACG CCAC.GTTTC
    501 TGAACGTGTC GTTTATCGTA TTCGCGCTGT TTTTCGTGCC GTATTTCGAT
    551 CCGCCCGTTA CCGCGCyGGC GTGGGCGGTC TTTGTCGGCG GCATTTTGCA
    601 ACTCGmTTC CAACTGCCCT GGCTGGCGAA ACTGGGCTTT TTGAAACTGC
    651 CCAAACTGAG TTTCAAAGAT GCGGCGGTCA ACCGCGTGAT GAAACAGATG
    701 GCGCCTGCgA TTTTgGGCGT GAgCGTGCGC CAGGTTTCTT TGGTGATCAA
    751 CACGATTTTc GCGTCTTATC TGCAATCGGG CAGCGTTTCA TGGATGTATT
    801 ACGCCGACCG CATGATGGAG CTGCCAGCG GCGTGCTGGG GGCGGCACTT
    851 GGTACGATTT TGCTGCCGAC TTTGTCCAAA CACTCGGCAA ACCaAGATAC
    901 GGaACAGTTT TCCGCCCTGC TCGACTGGGG TTTGCGCCTG TGCATGctgc
    951 TGACGCTGCC GCGGgcGGT GGA CTGGCGG TGTTGTCTGTT cCCgCtGGTG
    1001 CCGACGCTGT TTATGTACCG CGwATTTACG CTGTTTGACG CGCAGATGAC
    1051 GCAACACGCG CTGATTGCCT ATTCTTTCGG TTTAATCGGC TTAATCATGA
    1101 TTAAAGTGTT GGCACCCGGC TTCTATGCGC GGCAAAACAT CAAwAmGCCC
    1151 GTCAAAATCG CCATCTTCAC GCTCATCTGC mCGCAGTTGA TGAACCTTGs
    1201 CTTTayCGGC CCACTrrAAC rCaTcGGAC TTTGCTTGC CATCGGTCTG
    1251 GGCGCGTGTA TCAATGCCGG ATTGTTGTTT TACCTGTGTC GCAGACACGG
    1301 TATTTACCAA CCTGG.CAAG GGTGGGCAG CGTTCTT.AG CAAAAATGCT
    1351 GcTCTCGCTC GCCGTGA
  
```

This corresponds to the amino acid sequence <SEQ ID 114; ORF20>:

```

45      1 MNMLGALAKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAEKLPNL
      51 LRRVFAEGAF AQAFVPILAE YKETRSKEAX EAFIRHVAGM LSFVLVIVTA
    101 LGILAAPWVI YVSAPSAQD ADKFQLSIDL LRITFPYILL ISLSSFVGSV
      151 LNSYHKFGIP AFTPXFLNVS FIVEALFFVP YFDPPTAXA WAVEVGGILQ
    201 LXFQLPWLAK LGFLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQVSLVIN
    251 TIFASYLQSG SVSWMYADR MMELPSGVLG AALGTILLPT LSKHSANQDT
    301 EQFSALLDWG LRLCMLLTLP AAVGLAVLSF PLVATLFMYR XFTLFDAQMT
    351 QHALIAYSGF LIGLIMIKVL APGFYARQNI XXPVKIAIFT LICXQLMNLX
    401 FXGPLXXIGL SLAIGLGACI NAGLLFYLLR RHGIYQXPQG LGSVLXQKCC
    451 SRSP*
  
```

These sequences were elaborated, and the complete DNA sequence <SEQ ID 115> is:

```

55      1 ATGAATATGC TGGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTC
      51 GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTGCGCG
  
```

101 CGGGTATGGC GACGGATGCG TTTTGTGTCG CGTTCAAACCT GCCCAACCTG  
 151 CTTGCGCCGCG TGTGTCGCGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT  
 201 TTTGGCGGAA TACAAGGAAA CGCGTTCAAA AGAGGCGCGG GAGGCTTTTA  
 251 TCCGCCATGT GCGGGGATG CTGTCGTTTG TACTGGTTAT CGTTACCGCG  
 5 301 CTGGGCATAC TTGCCCGCGC TTGGGTGATT TATGTTTCCG CACCGGTTT  
 351 TGCCCAAGAT GCCGACAAAT TTCAGCTCTC CATCGATTG CTGCGGATTA  
 401 CGTTTCCTTA TATATATTG ATTCCCTGT CTTCAATTGT CGGCTCGGTA  
 451 CTCAATTCTT ATCATAAGTT CGGCATTCCG GCGTTTACGC CCACGTTTCT  
 10 501 GAACGTGTCG TTTATCGTAT TCGCGCTGTT TTTCGTGCGG TATTTCGATC  
 551 CGCCCGTTAC CGCGCTGGCG TGGGCGGTCT TTGTGCGCGG CATTTCGCAA  
 601 CTCGGCTTCC AACTGCCCTG GCTGGCGAAA CTGGGCTTTT TGAACCTGCC  
 651 CAAACTGAGT TTCAAAGATG CGGCGGTCAA CCGCGTGATG AAACAGATGG  
 701 CGCCTGCGAT TTTGGGCGTG AGCGTGGCGC AGGTTTCTTT GGTGATCAAC  
 751 ACGATTTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATGTATTA  
 15 801 CGCCGACCGC ATGATGGAGC TGCCGACGGG CGTGCTGGGG GCGGCACTCG  
 851 GTACGATTTT GCTGCCGACT TTGTCCAAAC ACTCGGCAAA CCAAGATACG  
 901 GAACAGTTT CCGCCCTGCT CGACTGGGGT TTGCGCTGT GCATGCTGCT  
 951 GACGCTGCCG GCGGCGGTG GACTGGCGGT GTTGTGCTTC CCGCTGGTGG  
 1001 CGACGCTGTT TATGTACCGC GAATTTACGC TGTGTTGACG GCAGATGACG  
 20 1051 CAACACGCGC TGATTGCCTA TTCTTCGGT TTAATCGGCT TAATCATGAT  
 1101 TAAAGTGTG GCACCGGCT TCTATGCGCG GCAAAACATC AAAACGCCCG  
 1151 TCAAATCGC CATCTTCACG CTCATCTGCA CGCAGTTGAT GAACCTTGCC  
 1201 TTTATCGGCC CACTGAAACA CGTCGGACTT TCGCTTGCCA TCGTCTGGG  
 25 1251 CGCGTGTATC AATGCCGGAT TGTGTTTCTA CCTGTTGCGC AGACACGGTA  
 1301 TTTACCAACC TGGCAAGGGT TGGGCAAGCT TCTTAGCAAA AATGCTGCTC  
 1351 TCGCTCGCCG TGATGTGCGG CGGACTGTGG GCAGCGCAGG CTTACCTGCC  
 1401 GTTTGAATGG GCGCAGCCG GCGGAATCGG GAAAGCGGG CAGCTCTGCA  
 1451 TCCTGATTGC CGTCGCGGC GGAAGTGTATT TCGCATCACT GCGGCTTTG  
 1501 GGCTTCCGTC CGCGCCATTT CAAACGCGTG GAAACTGA

30 This corresponds to the amino acid sequence <SEQ ID 116; ORF20-1>:

1 MNMLGALAKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAFKLPNL  
 51 LRRVFAEGAF AQAFVPILAE YKETRSKEAA EAFIRHVAGM LSFVLVIVTA  
 101 LGILAAPWVI YVSAPGFAQD ADKFQLSIDL LRITFPYILL ISLSFVGSV  
 151 LNSYHKFGIP AFTPTFLNVS FIVFALFFVP YFDPPTALA WAVFVGGILO  
 35 201 LGFQLPWLAK LGFLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQVSLVIN  
 251 TIFASYLQSG SVSWMYYADR MMELPSGVLG AALGTILLPT LSKHSANQDT  
 301 EQFSALLDWG LRLCMLLTL AAVGLAVLSF PLVATLEMYR EFTLFDAQMT  
 351 QHALIAYSFG LIGLIMIKVL APGFYARQNI KTPVKIAIFT LICTQLMNL  
 401 FIGPLKHVGL SLAIGLGACI NAGLLFYLLR RHGIYQPGKG WAAFLAKMLL  
 40 451 SLAVMCGGLW AAQAYLPFEW AHAGGMRKAG QLCILIAVGG GLYFASLAAL  
 501 GFRPRHFKRV EN\*

Computer analysis of this amino acid sequence gave the following results:

Homology with the MviN virulence factor of *S. typhimurium* (accession number P37169)

ORF20 and MviN proteins show 63% aa identity in 440aa overlap:

45 Orf20 1 MNMLGALAKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAFKLPNL LRRVFAEGAF 60  
 MN+L +LA V S+TM SRVLGF RD ++AR FGAGMATDA FFVAFKLPNL LRR+FAEGAF  
 MviN 14 MNLLKSLAAVSSMTMF SRVLGFARDAIVARIFGAGMATDA FFVAFKLPNL LRRIFAEGAF 73  
 50 Orf20 61 AQAFVPILAEYKETRSKEAXEAFIRHVAGM LSFVLVIVTALGILAAPWVIYVSAPGFAQD 120  
 +QAFVPILAEYK + +EA F+ +V+G+L+ L +VT G+LAAPWVI V+AP FA  
 MviN 74 SQAFVPILAEYKSKQGEATRIFVAYVSGLLTLALAVVTVAGMLAAPWVIMVTAPGFADT 133  
 55 Orf20 121 ADKFQLSIDLLRITFPYILLISLSFVGSV LNSYHKFGIP AFTPTFLNVS FIVFALFFVP 180  
 ADKF L+ LLRITFPYILLISL+S VG++LN+++F IPAF P FLN+S I FALF P  
 MviN 134 ADKFALTQLLRITFPYILLISLASLVGAILNTWNRF SIPAFAPTFLNISMIGFALFAAP 193  
 60 Orf20 181 YFDPPTAXAWAVFVGGILQLXFLPWLAKLGFLKLPKLSFKDAAVNRVM KQMAPAILGV 240  
 YF+PPV A AWAV VGG+LQL +QLP+L K+G L LP+++F+D RV+KQM PAILGV  
 MviN 194 YFNPPVLALAWAVTVGGVLQLVYQLPYLKKIGMLVLPRI NFRDTGAMRVVKQMPAILGV 253  
 Orf20 241 SVAQVSLVINTIFASYLQSGSVSWMYYADRM MELPSGVLGAALGTILLPTLSKHSANQDT 300  
 SV+Q+SL+INTIFAS+L SGSVSWMYYADR+ME PSGVLG ALGTILLP+LSK A+ +  
 MviN 254 SVSQISLIINTIFASFLASGSVSWMYADRLMEF PSGVLGVALGTILLPSLSKSFASGNH 313

-117-

Orf20 301 EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXFTLFDAQMTQHAIAYSFG 360  
 +++ L+DWGLRLC LL LP+AV L +L+ PL +LF Y FT FDA MTQ ALIAYS G  
 MviN 314 DEYCRIMDWGLRLCFLALPSAVALGILAKPLTVSLFQYGKFTAFDAAMTQRALIAYSVG 373

5 Orf20 361 LIGLIMIKVLAPGFYARONIXXPVKIAIFTLICXQLMNLXFXXXXXXXXXXXXXXXXXXCI 420  
 LIGLI++KVLAPGFY+RQ+I PVKIAI TLI QLMNL F C+  
 MviN 374 LIGLIVVKVLAPGFYSRQDIKTPVKIAIVTLIMTQLMNLAFIGPLKHAGLSLSIGLAACL 433

10 Orf20 421 NAGLLFYLLRRHGIYQXPQG 440  
 NA LL++ LR+ I+ P G  
 MviN 434 NASLLYWQLRKQNIPTPQG 453

### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF20 shows 93.5% identity over a 447aa overlap with an ORF (ORF20a) from strain A of *N.*

*meningitidis*:

	10	20	30	40	50	60
orf20.pep	MNMLGALAKVGS	SLTMVSRVLG	FVRDVTIARA	FGAGMATDAFF	VAFKLPNLLRR	VFAEGAF
orf20a	MNMLGALVKVGS	SLTMVSRVLG	FVRDVTIARA	FGAGMATDAFF	VAFKLPNLLRR	VFAEGAF
	70	80	90	100	110	120
orf20.pep	AQAFVPILA	EYKETR	SKEAXEAFIR	HVAGMLS	FVLVIVTAL	GILAAPWVIYVSAPSFAQD
orf20a	AQAFVPILA	EYKETR	SKEATEAFIR	HVAGMLS	FVLVIVTAL	GILAAPWVIYVSAPGFAKD
	130	140	150	160	170	180
orf20.pep	ADKFQLSID	LLRITFPY	ILLISLS	SFVGSVLNS	YHKFGIPAFT	PKFLNVSFIVFALEFFVP
orf20a	ADKFQLSID	LLRITFPY	ILLISLS	SFVGSVLNS	YHKFSIPAFT	PTFLNVSFIVFALEFFVP
	190	200	210	220	230	240
orf20.pep	YFDPPVTAX	AWAVFVG	GILQLX	FQLPWLAK	LGFLKLPKLS	FKDAAVNRVMKQMAPAILGV
orf20a	YFDPPVTAL	AWAVFVG	GILQLG	FQLPWLAK	LGFLKLPKLS	FKDAAVNRVMKQMAPAILGV
	250	260	270	280	290	300
orf20.pep	SVAQVSLV	INTIFASY	LQSGSVSW	MYADRMEL	PSGVLGAAL	GTILLPTLSKHSAHQDT
orf20a	SVAQISLV	INTIFASY	LQSGSVSW	MYADRMEL	PGGVLGAAL	GTILLPTLSKHSAHQDT
	310	320	330	340	350	360
orf20.pep	EQFSALLDW	GLRLCML	LTLPAAVGL	AVLSFPLVAT	LFMYRXFTLF	DAQMTQHAIAYSFG
orf20a	EQFSALLDW	GLRLCML	LTLPAAVGL	AVLSFPLVAT	LFMYREFTLF	DAQMTQHAIAYSFG
	370	380	390	400	410	420
orf20.pep	LIGLIMIKV	LAPGFYAR	ONIXXPVK	IAIFTLICX	QLMNLXFX	GPLXXIGLSLAIGLGACI
orf20a	LIGLIMIKV	LAPGFYAR	ONIKTPVK	IAIFTLIC	QLMNLAFI	GPLKHVGLSLAIGLGACI
	430	440	450			
orf20.pep	NAGLLFYLL	RRHGIYQ	XPQGLG	SVLXQK	CCSRSPX	
orf20a	NAGLLFYLL	RRHGIYQ	PGKGWAA	FLAKMLLS	LAVMGGGL	YAAQIWLFPDWAHAGGMQKAA
	460	470	480			

The complete length ORF20a nucleotide sequence <SEQ ID 117> is:

1 ATGAATATGC TGGGAGCTTT GGTAAAAGTC GGCAGCCTGA CGATGGTGTC  
 51 GCGCGTTTGG GATTGTTGTC GCGATACGGT CATTGCGCGC GCATTGCGCG  
 101 CAGGCATGGC GACGGATGCG TTCTTTGTCG CGTTCAAACT GCCCAACCTG

-118-

5  
10  
15  
20  
25

```

151 CTTCGCCGCG TGTTTGCGGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT
201 TTTGGCGGAA TATAAGGAAA CGCGTTCTAA AGAGGCGACG GAGGCTTTTA
251 TCCGCCATGT GCGGGGATG CTGTCGTTTG TACTGGTCAT CGTTACCGCG
301 CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTTCCG CACCCGGTTT
351 TGCCAAAGAT GCCGACAAAT TTCAGCTCTC TATCGATTG CTGCGGATTA
401 CGTTTCCTTA TATCTTATTG ATTTCACTTT CCTCTTTTGT CGGCTCGGTA
451 CTCAATTCCT ATCATAAATT CAGCATTCCT GCGTTTACGC CCACGTTTCT
501 GAACGTGTCG TTTATCGTAT TCGCGCTGTT TTTCGTGCGG TATTTCGATC
551 CTCCCGTTAC CGCGCTGGCT TGGGCGGTTT TTGTGCGCGG CATTTTGCAA
601 CTCGGCTTCC AACTGCCCTG GCTGGCGAAA CTGGGTTTTT TGAAACTGCC
651 CAAACTGAGT TTCAAAGATG CCGCGGTCAA CCGCGTGATG AAACAGATGG
701 CGCGTGCAT TTTGGGCGTG AGCGTGCGC AGATTTCTTT GGTGATCAAC
751 ACGATTTTCG CGTCTTATCT GCAATCGGCG AGCGTTTCAT GGATGTATTA
801 CGCCGACCGC ATGATGGAAC TGCCCGCGCG CGTGCTGGGG GCGGCACTCG
851 GTACGATTTT GCTGCCGACT TTGTCCAAAC ACTCGGCAAA CCAAGATACG
901 GAACAGTTTT CCGCCCTGCT CGACTGGGGT TTGCGCNTGT GCATGCTGCT
951 GACGCTGCCG GCGGCGGTCG GAATGGCGGT GTTGTGCTTC CCGCTGGTGG
1001 CAACCTTGTT TATGTACCGA GAATTCACGC TGTGTGACGC GCAGATGACG
1051 CAACACGCGC TGATTGCCTA TTCTTTCGGT TTAATCGGTT TAATCATGAT
1101 TAAAGTGTTG GCGCCCGGCT TTTATGCGCG GCAAAACATC AAAACGCCCCG
1151 TCAAAATCGC CATCTTCACG CTCATTTGCA CGCAGTTGAT GAACCTTGCC
1201 TTTATCGGCC CACTGAAACA CGTCGGACTT TCGCTTGCCA TCGGTCTGGG
1251 CGCGTGATC AATGCCGAT TGTTGTTTTA CCTGTTGCGC AGACACGGTA
1301 TTTACCAACC TGGCAAGGGT TGGGCAGCGT TCTTGGCAAA AATGCTGCTC
1351 TCGCTCGCCG TGATGGGAGG CGGCTGTAT GCCGCCAAA TCTGGCTGCC
1401 GTTCGACTGG GCACACGCCG GCGGAATGCA AAAGGCCGCC CGGCTCTTCA
1451 TCCTGATTGC CGTCGGCGGC GGAAGTATT TCGCATCACT GCGCGCTTTG
1501 GGCTTCCGTC CGCGCCATT CAAACGCGTG GAAAGCTGA

```

This encodes a protein having amino acid sequence <SEQ ID 118>:

30  
35  
40

```

1 MNMLGALVKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAFKLPNL
51 LRRVFAEGAF AQAFVPILAE YKETRSKEAT EAFIRHVAGM LSEFLVIVTA
101 LGILAAPWVI YVSAPGFAKD ADKFQLSIDL LRITFPYILL ISLSSFVGSV
151 LNSYHKFSIP AFTPTFLNVS FIVEFAFFVP YFDPPTALAW WAVEFGGILQ
201 LGFQLPWLAK LGFLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQISLVIN
251 TIFASYLQSG SVSWMYADR MMELPGGVLG AALGTILLPT LSKHSANQDT
301 EQFSALLDWG LRXCMLTLP AAVGMAVLSF PLVATLFMYR EFTLFDAQMT
351 QHALIAYSFG LIGLIMIKVL APGFYARQNI KTPVKIAIFT LICTQLMNLA
401 FIGPLKHVGL SLAIGLGACI NAGLLFYLLR RHGIYQPGKG WAAFLAKMLL
451 SLAVMGGGLY AAQIWLFPDW AHAGGMQKAA RLFILIAVGG GLYFASLAAL
501 GFRPRHFKRV ES*

```

ORF20a and ORF20-1 show 96.5% identity in 512 aa overlap:

45  
50  
55  
60  
65

```

              10      20      30      40      50      60
orf20a.pep  MNMLGALVKVGS LTMVSRVLGFVRDVIARA FGAAGMATDAFFVAFKLPNLLRRVFAEGAF
              10      20      30      40      50      60
orf20-1     MNMLGALAKVGS LTMVSRVLGFVRDVIARA FGAAGMATDAFFVAFKLPNLLRRVFAEGAF
              70      80      90      100     110     120
orf20a.pep  AQAFVPILAEYKETRSKEATEAFIRHVAGM LSEFLVIVTALGILAAPWVIYVSAPGFAKD
              70      80      90      100     110     120
orf20-1     AQAFVPILAEYKETRSKEAAEAFIRHVAGM LSEFLVIVTALGILAAPWVIYVSAPGFAQD
              130     140     150     160     170     180
orf20a.pep  ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFSIPAFPTPTFLNVSFIVFALEFFVP
              130     140     150     160     170     180
orf20-1     ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFIPAFPTPTFLNVSFIVFALEFFVP
              190     200     210     220     230     240
orf20a.pep  YFDPPTALAWAVEFGGILQIGFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV
              190     200     210     220     230     240
orf20-1     YFDPPTALAWAVEFGGILQIGFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV
              250     260     270     280     290     300
orf20a.pep  SVAQISLVINTIFASYLQSGSVSWMYADRM MELPGGVLGAALGTILLPTLSKHSANQDT

```

-119-

	orff20-1	:                                :
		SVAQVSLVINTIFASYLQSGSVSWMYADRMELPSGVLGAALGTILLPTLSKHSANQDT
		250 260 270 280 290 300
5	orff20a.pep	310 320 330 340 350 360
		EQFSALLDWGLRXCMLLTLPAAVGMVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG
	orff20-1	
		EQFSALLDWGLRXCMLLTLPAAVGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG
10		310 320 330 340 350 360
	orff20a.pep	370 380 390 400 410 420
		LIGLIMIKVLAPGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHVGLSLAIGLGACI
	orff20-1	
15		LIGLIMIKVLAPGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHVGLSLAIGLGACI
		370 380 390 400 410 420
	orff20a.pep	430 440 450 460 470 480
		NAGLLFYLLRRHGIYQPGKGWAAFLAKMLLSLAVMGGGLYAAQIWLPPFWAHAGGMQKAA
20	orff20-1	
		NAGLLFYLLRRHGIYQPGKGWAAFLAKMLLSLAVMCGGLWAAQAYLPFEWAHAGGMKAG
		430 440 450 460 470 480
	orff20a.pep	490 500 510
		RLFILIAVGGGLYFASLAALGFRPRHEKRVESX
25	orff20-1	:       :     :     :     :     :     :     :
		QLCILIAVGGGLYFASLAALGFRPRHEKRVENX
		490 500 510

Homology with a predicted ORF from *N.gonorrhoeae*

30 ORF20 shows 92.1% identity over a 454aa overlap with a predicted ORF (ORF20ng) from *N. gonorrhoeae*:

	orff20.pep	MNMLGALAKVGS LTMVSRVLGFVRD TVIARA FGAGMATDAFFVAFKLPNLLRRVFAEGAF	60
	orff20ng	MNMLGALAKVGS LTMVSRVLGFVRD TVIARA FGAGMATDAFFVAFKLPNLLRRVFAEGAF	60
35	orff20.pep	AQAFVPILA EYK ETRSK EAXEAFIRHVAGMLS FVLVIVTALGILAAPWVIYVSAPSFAQD	120
	orff20ng	AQAFVPILA EYK ETRSK EATEAFIRHVAGMLS FVLIVVTALGILAAPWVIYVSAPGFTKD	120
40	orff20.pep	ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIPAFTPXFLNVSFIVFALFFVP	180
	orff20ng	ADKFQLSISLLRITFPYILLISLSSFVGSILNSYHKFGIPAFTPTFLNISFIVFALFFVP	180
45	orff20.pep	YFDPPVTAXAWAVFVGILQLXFLPWLAKLGLKLPKLSFKDAAVNRVMKQMAPAILGV	240
	orff20ng	YFDPPVTALAWAVFVGILQLGFQLPWLAKLGLKLPKLNFKDAAVNRVMKQMAPAILGV	240
	orff20.pep	SVAQVSLVINTIFASYLQSGSVSWMYADRMELPSGVLGAALGTILLPTLSKHSANQDT	300
50	orff20ng	SVAQISLVINTIFASYLQSGSVSWMYADRMELPGGVLGAALGTILLPTLSKHSANQDT	300
	orff20.pep	EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXFTLFDAQMTQHAIAYSFG	360
	orff20ng	EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG	360
55	orff20.pep	LIGLIMIKVLAPGFYARQNIKTPVKIAIFTLICTQLMNLXFXGPLXXIGLSLAIGLGACI	420
	orff20ng	LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI	420
60	orff20.pep	NAGLLFYLLRRHGIYQPXQGLGSVLXQKCCSRSP	454
	orff20ng	NAGLLFFLFRKHGIYRPGQGLQPSWRKCCSRSP	454

An ORF20ng nucleotide sequence <SEQ ID 119> was predicted to encode a protein having amino acid sequence <SEQ ID 120>:

-120-

1 MNMLGALAKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAFKLPNL  
 51 LRRVFAEGAF AQAFVPILAE YKETRSKEAT EAFIRHVAGM LSFVLIVVTA  
 101 LGILAAPWVI YVSAPGFTKD ADKFQLSISL LRITFPYILL ISLSSFVSGSI  
 151 LNSYHKFGIP AFTPTFLNIS FIVFALFFVP YFDPPTALA WAVFVGILQ  
 5 L 201 LGFQLPWLAK LGFLKLPKLN FKDAAVNRVM KQMAPAILGV SVAQISLVIN  
 251 TIFASYLQSG SVSWMYADR MMELPGGVLG AALGTILLPT LSKHSANQDT  
 301 EQFSALLDWG LRLCMLLTLP AAAGLAVLSF PLVATLFMYR EFTLFDAQMT  
 351 QHALIAYSFG LIGLIMIKVL ASGFYARQNI KTPVKIAIFT LICTQLMNLA  
 10 401 FIGPLKHAGL SLAIGLGACI NAGLLFFLFR KHGIYRPGQG LGQPSWRKCC  
 451 SRSP\*

Further DNA sequence analysis revealed the following DNA sequence <SEQ ID 121>:

1 ATGAATATGC TTGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTCT  
 51 GCGCGTTTTG GGATTGTGTC GCGATACGGT CATTGCGCGG GCATTGCGCG  
 15 101 CGGGTATGGC GACGGATGCG TTTTGTGTCG CGTTCAACT GCCCAACCTG  
 151 CTTCGCGCGG TGTTCGCGGA GGGGCGGTTT GCCCAACGCT TTGTGCGCG  
 201 TTTGGCGGAA TATAAGGAAA CGCGTTCTAA AGAGGCGAag gAGGCTTTTA  
 251 TCCGCCACGt tgcgggAatg CTGTCGTTTG TGCTGATcgt cGttacCGCG  
 301 CTGGGCATAC TTGCCGCGcc tTGGGTGATT TATGTTtccg CgcccGGCTT  
 351 TACCAAGAC GCGGACAAAT TCCAACTTTC CATCAGCCTG CTGCGGATTA  
 20 401 CGTTTCCTTA TATATTATTG ATTTCTTTGT CTTCTTTTGT CGGCTCGATA  
 451 CTAATTCCT ACCATAAGTT CGGCATTCCC GCGTTTACGC CCACGTTTTT  
 501 AAACATCTCT TTTATCGTAT TCGCACTGTT TTTCGTGCCG TATTTCGATC  
 551 CGCCGTTTAC CGCGTGGCG TGGGCGGTTT TTGTCGCGCG TATTTGTCAG  
 601 CTCGGTTTCC AACTGCGGTG GCTGGCGAAA CTGGGCTTTT TGAAACTGCC  
 25 651 CAACTGAAT TTCAAGATG CGGCGGTCAA CCGCGTCATG AACAGATGG  
 701 CGCTGCGAT TTTGGGCGTG agcgTGGCGC AAATTTCTTT GgttATCAAC  
 751 ACGATTTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATGTatta  
 801 cGCCGACCGC ATGATGGAGc tgcgccGGGG CGTGCTGGGG GCTGCACTCG  
 851 GTACAATTTT GCTGCCGACT TTGTCCAAAC ACTCGGCAA CCAAGATACG  
 30 901 GAACAGTTTT CCGCCCTGCT CGACTGGGGT TTGCGCCTGT GCATGCTGCT  
 951 GACGCTGCCG GCGGCGGccg GACTGGCGGT ATTGTCGTTT CCGCTGGTGG  
 1001 CGAGCTGTT TATGTACCGA GAATTCACGC TGTTTGACGC ACAATGACG  
 1051 CAACACGCGC TGATTGCCTA TTCTTTCGGT TTAATCGGTT TAATTATGAT  
 35 1101 TAAAGTGTG GCATCCGCT TTTATGCGCG GCAAAACATC AAAACGCCCCG  
 1151 TCAAATCGC CATCTTCACG CTCATCTGCA CGCAGTTGAT GAACCTCGCC  
 1201 TTTATCGGTC CGTTGAAACA CGCGGGCTT TCGCTCGCCA TCGGCTGGG  
 1251 CGCTGTCATC AACGCCGAT TGTGTTCTT CTGTTGCGC AAACACGGTA  
 1301 TTTACCGGCC cggcaggggt tgggcggcgt TCTTGGCGAA AATGCTGCTC  
 1351 GCGCTCGCCG TGATGTGCGG CCGACTGTGG GCGGCGCAGG CTTGCTGCC  
 40 1401 TTTCGAATGG GCGCACGCCG GCGGAATGCG GAAAGCGGGG CAGCTCTGCA  
 1451 TCCTGATTGC CGTCGGCGGC GGACTGTATT TCGCATCTCT GGCGGCTTTG  
 1501 GGCTTCGTC CGGCCATTT CAAACGCGTG GAAAGCTGA

This encodes the following amino acid sequence <SEQ ID 122; ORF20ng-1>:

1 MNMLGALAKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAFKLPNL  
 45 51 LRRVFAEGAF AQAFVPILAE YKETRSKEAT EAFIRHVAGM LSFVLIVVTA  
 101 LGILAAPWVI YVSAPGFTKD ADKFQLSISL LRITFPYILL ISLSSFVSGSI  
 151 LNSYHKFGIP AFTPTFLNIS FIVFALFFVP YFDPPTALA WAVFVGILQ  
 50 201 LGFQLPWLAK LGFLKLPKLN FKDAAVNRVM KQMAPAILGV SVAQISLVIN  
 251 TIFASYLQSG SVSWMYADR MMELRRGVLG AALGTILLPT LSKHSANQDT  
 301 EQFSALLDWG LRLCMLLTLP AAAGLAVLSF PLVATLFMYR EFTLFDAQMT  
 351 QHALIAYSFG LIGLIMIKVL ASGFYARQNI KTPVKIAIFT LICTQLMNLA  
 401 FIGPLKHAGL SLAIGLGACI NAGLLFFLLR KHGIYRPGRG WAAFLAKMLL  
 451 ALAVMCGGLW AAQACLPEFW AHAGGMRKAG QLCILIAVGG GLYFASLAAL  
 501 GFRPRHFKRV ES\*

ORF20ng-1 and ORF20-1 show 95.7% identity in 512 aa overlap:

	10	20	30	40	50	60
orf20-1.pep	MNMLGALAKV	GSLTMVSRVL	GFVRDVIAR	AFGAGMATDA	FFVAFKLPNL	LRRVFAEGAF
orf20ng-1	MNMLGALAKV	GSLTMVSRVL	GFVRDVIAR	AFGAGMATDA	FFVAFKLPNL	LRRVFAEGAF
	10	20	30	40	50	60
	70	80	90	100	110	120
orf20-1.pep	AQAFVPILAE	YKETRSKEA	EAFIRHVAGM	LSFVLIVVTA	LGILAAPWVI	YVSAPGFAQD
orf20ng-1	AQAFVPILAE	YKETRSKEA	EAFIRHVAGM	LSFVLIVVTA	LGILAAPWVI	YVSAPGFTKD



-121-

		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf20-1.pep	ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIPAFPTPTFLNVSFIVFALFFVP					
	orf20ng-1	ADKFQLSISLLRITFPYILLISLSSFVGSILNSYHKFGIPAFPTPTFLNISFIVFALFFVP					
		130	140	150	160	170	180
10	orf20-1.pep	YFDPPTALAWAVFVGGILQLGFQLPWLAKLGLKLPKLSFKDAAVNRVMKQMAPAILGV					
	orf20ng-1	YFDPPTALAWAVFVGGILQLGFQLPWLAKLGLKLPKLNFKDAAVNRVMKQMAPAILGV					
		190	200	210	220	230	240
15	orf20-1.pep	SVAQVSLVINTIFASYLQSGSVSWMYADRMELPSGVLGAALGTILLPTLSKHSANQDT					
	orf20ng-1	SVAQISLVINTIFASYLQSGSVSWMYADRMELRRGVLGAALGTILLPTLSKHSANQDT					
		250	260	270	280	290	300
20	orf20-1.pep	EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSEF					
	orf20ng-1	EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSEF					
25		310	320	330	340	350	360
	orf20-1.pep	LIGLIMIKVLAPGFYARONIKTPVKIAIFTLICTQLMNLAFIGPLKHVGLSLAIGLGACI					
	orf20ng-1	LIGLIMIKVLASGFYARONIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI					
		370	380	390	400	410	420
30	orf20-1.pep	NAGLLFYLLRRHGIYQPGKGWAAFLAKMLLSLAVMCGGLWAAQAYLPFEWAHAGGMRKAG					
	orf20ng-1	NAGLLFFLLRKHGIYRPGRGWAAFLAKMLLALAVMCGGLWAAQACLPFEWAHAGGMRKAG					
		430	440	450	460	470	480
35	orf20-1.pep	QLCILIAVGGGLYFASLAALGFRPRHFKRVENX					
	orf20ng-1	QLCILIAVGGGLYFASLAALGFRPRHFKRVESX					
		490	500	510			

In addition, ORF20ng-1 shows significant homology with a virulence factor of *S.typhimurium*:

45	sp P37169 MVIN_SALTY VIRULENCE FACTOR MVIN pir  S40271 mviN protein - Salmonella typhimurium gi 438252 (226133) mviB gene product [Salmonella typhimurium] gnl PID d1005521 (D25292) ORF2 [Salmonella typhimurium] Length = 524 Score = 1573 (750.1 bits), Expect = 1.1e-220, Sum P(2) = 1.1e-220 Identities = 309/467 (66%), Positives = 368/467 (78%)
50	Query: 1 MNMLGALAKVGSMTVSRVLGFVRDVIARAFAFGAGMATDAFFVAFKLPNLLRRVFAEGAF 60 MN+L +LA V S+TM SRVLGF RD ++AR FGAGMATDAFFVAFKLPNLLRR+FAEGAF Sbjct: 14 MNLLKSLAAVSSMTMSRVLGFARDIAVARIFGAGMATDAFFVAFKLPNLLRRIFAEGAF 73
55	Query: 61 AQAFVPILAIEYKETSKEATEAFIRHVAGMLSEVLIVVTALGILAAPWVIYVSAPGFTKD 120 +AQAFVPILAIEYK + +EAT F+ +V+G+L+ L VVT G+LAAPWVI V+APGF Sbjct: 74 SQAFVPILAIEYKSKQGEETRIFVAYVSGLLTLALAVVTAGMLAAPWVIMVTAPGFADT 133
60	Query: 121 ADKFQLSISLLRITFPYILLISLSSFVGSILNSYHKFGIPAFPTPTFLNISFIVFALFFVP 180 ADKF L+ LLRITFPYILLISL+S VG+ILN++++F IPAF PTFLNIS I FALF P Sbjct: 134 ADKFALTQLLRITFPYILLISLASLVGAILNTWNRFSIPAFAPTFLNISMIGFALFAAP 193
65	Query: 181 YFDPPTALAWAVFVGGILQLGFQLPWLAKLGLKLPKLNFKDAAVNRVMKQMAPAILGV 240 YF+PPV ALAWAV VGG+LQL +QLP+L K+G L LP++NF+D RV+KQM PAILGV Sbjct: 194 YFNPPVLALAWAVTVGGVLQVYQLPYLKKIGMLVLPRIINFRDGTAMRVVKQMGPAAILGV 253
70	Query: 241 SVAQISLVINTIFASYLQSGSVSWMYADRMELRRGVLGAALGTILLPTLSKHSANQDT 300 SV+QISL+INTIFAS+L SGSVSWMYADR+ME GVLG ALGTILLP+LSK A+ + Sbjct: 254 SVSQISLIINTIFASFLASGSVSWMYADRMEFPSGVLGVALGTILLPSLSKSFASGNH 313

-122-

Query: 301 EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSEF 360  
 +++ L+DWGLRLC LL LP+A L +L+ PL +LF Y +FT FDA MTQ ALIAYS G  
 Sbjct: 314 DEYCRLMDWGLRLCFLALPSAVALGILAKPLTVSLFQYGKFTAFDAAMTORALIAYSVG 373

5 Query: 361 LIGLIMIKVLASGFYARONIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI 420  
 LIGLI++KVLG GFY+RQ+IKTPVKIAI TLI TQLMNLAFIGPLKHAGLSL+IGL AC+  
 Sbjct: 374 LIGLIVKVLAPGFYSRQDIKTPVKIAIVTLIMTQLMNLAFIGPLKHAGLSLSIGLAACL 433

10 Query: 421 NAGLLFFLLRKHGIIYRPGRGWXXXXXXXXXXXXXVMCGGLWAAQACLP 467  
 NA LL++ LRK I+ P GW VM L+ +P  
 Sbjct: 434 NASLLYWQLRKQNIPTPQPGWMWFLMRLIISVLVMAAVLFGVLHIMP 480

Score = 70 (33.4 bits), Expect = 1.1e-220, Sum P(2) = 1.1e-220  
 Identities = 14/41 (34%), Positives = 23/41 (56%)

15 Query: 469 EWAHAGGMRKAGQLCILIAVGGGLYFASLAALGFRPRHEKR 509  
 EW+ + + +L ++ G YFA+LA LGF+ + F R  
 Sbjct: 481 EWSQGSMLWRLRLMAVVIAGIAAYFAALAVLGFKVKEFVR 521

- 20 Based on this analysis, including the homology with a virulence factor from *S.typhimurium*, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 15

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 123>:

25 1 atGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA  
 51 GCAAGCCGTT tACGACGGCC CGGCCATTAC CGAAGTCGCG TTGCTTGCGG  
 101 AAGAATATGC CGGTATGCGC CCCTCGATGA AAGTCAAGGA AGGCGATGCC  
 151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGGT  
 201 GTTTACTGCG CCGGCTTCAG GcAAAATCGC CGCGATTAC CGTGGCGAAA  
 30 251 AGCGCTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC  
 301 GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAACCTAA GCGGCGAAGA  
 351 AGTGCGCCGC AACCTGATCC AATCCGTTT GTGGACTGCG CTGCGCACCC  
 401 GTCCGTTTCA CAAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC  
 451 GTCAATGCGA tGACACCAA TCCG..

- 35 This corresponds to the amino acid sequence <SEQ ID 124; ORF22>:

1 MIKIKKGLNL PIAGRPEQAV YDGPATEVA LLGEEYAGMR PSMKVKEGDA  
 51 VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLSQV VIAVEXNDEI  
 101 EFERYAPEAL ANLSGEEVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF  
 151 VNAMDTPNP..

- 40 Further work revealed the complete nucleotide sequence <SEQ ID 125>:

1 ATGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA  
 51 GCAAGCCGTT TACGACGGCC CGGCCATTAC CGAAGTCGCG TTGCTTGCGG  
 101 AAGAATATGC CGGTATGCGC CCCTCGATGA AAGTCAAGGA AGGCGATGCC  
 151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGGT  
 45 201 GTTTACTGCG CCGGCTTCAG GCAAATCGC CGCGATTAC CGTGGCGAAA  
 251 AGCGCTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC  
 301 GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAACCTAA GCGGCGAAGA  
 351 AGTGCGCCGC AACCTGATCC AATCCGTTT GTGGACTGCG CTGCGCACCC  
 401 GTCCGTTTCA CAAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC  
 50 451 GTCAATGCGA TGGACACCAA TCCGCTGGCT GCCGACCCTA CGGTCAATTAT  
 501 CAAAGAAGCC GCCGAGGATT TCAAACGCGG CCTGTTGGTA TTGAGCCGTT  
 551 TGACCGAACG CAAAATCCAT GTTTGTAAGG CAGCTGGCGC AGACGTGCCG  
 601 TCTGAAAATG CTGCCAACAT CGAAACACAT GAATTCGGCG GCCCGCATCC  
 651 TGCCGGTTTG AGTGGCACCG ACATTCAATT CATCGAGCCG GTCGGCGCGA  
 701 ATAAAGACGT GTGGACCATC AATTATCAAG ATGTAATTAC CATTGGCCGT  
 751 TTGTTTGCAA CAGGCCGTCT GAACACCGAG CGCGTGATTG CCCTAGGTGG  
 801 TTCTCAAGTC AACAAACCGC GCCTCTTGGC TACCGTTTGG GGTGCGAAG  
 851 TATCGCAAT TACTGCGGGC GAATTGGTTG ACACAGACAA CCGCGTGATT  
 901 TCCGGTTCGG TATTGAACGG CGCGATTACA CAAGGCGCGC ACGATTATTT

5  
 951 GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG  
 1001 AGCTGTTTCGG CTGGGTTGCG CCGCAGCCGG ACAAATACTC CATCACGCGT  
 1051 ACAACCCTCG GCCATTTCCT GAAAAACAAA CTCTTCAAGT TCAACACAGC  
 1101 CGTCAACGGC GGCAGCCGCG CCATGGTGCC GATTGGTACT TACGAGCGCG  
 1151 TGATGCCCTT GGATATCCTG CCCACCCTGC TTTTGC CGCA TTTAATCGTC  
 1201 GGCATACCG ACAGCGCGCA GGCATTGGGT TGCTTGAAT TGGACGAAGA  
 1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATACGGCC  
 1301 CGCTGTTGCG CAAAGTGCTG GAAACCATTG AGAAGGAAGG CTGA

This corresponds to the amino acid sequence <SEQ ID 126; ORF22-1>:

10  
 1 MIKIKKGLNL PIAGRPEQAV YDGPATEVA LLGEEYAGMR PSMKVKEGDA  
 51 VKKGQVLFED KKNPGVVETA PASGKIAAIH RGEKRVLSV VIAVEGNDEI  
 101 EFERYAPEAL ANLSGEEVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF  
 151 VNAMDNTNPLA ADPTVIIKEA AEDEFKRLLV LSRLTERKIH VCKAAGADVP  
 201 SENAAINIETH EFGGPHFAGL SGTHIHFIET VGANKTVWTI NYQDVITIGR  
 15  
 251 LFATGRLNTE RVIALGGSQV NKPRLRLTVL GAKVSQITAG ELVDTDNRVI  
 301 SGSVLNGAIT QGAHDYLGRI HNQISVIEEG RSKELFGWVA PQPDKYSITR  
 351 TTLGHFLKNK LFKFTAVNG GDRAMVPIGT YERVMPLDIL PTLLLRDLIV  
 401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EYGPLLRKVL ETIEKEG\*

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 127>:

20  
 1 ATGATTAAAA TCAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA  
 51 GCAGTCATT TATGACGGG CCGTCATTAC CGAAGTCGCG TTGCTTGGCG  
 101 AAGAATATGC CGGTATGCGC CCCTNGATGA AAGTCAAGGA AGGCGATGCC  
 151 GTCAAAAAGG GCCAAGTGCT GTTGAAGAC AAAAAGNATC CGGGCGTGGT  
 201 GTTTACCGCG CCNGTTTCAG GCAAAATCGC CGCCATCCAT CGCGGCGAAA  
 25  
 251 AGCGCGTACT TCAGTCGGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC  
 301 GAGTTCGAAC GCTACGCGCC CGAAGCGTTG GCAAACCTAA GCGGCGGANGA  
 351 ANTNNNGNNGC AATCTGATCC AATCCGGTTT GTGGACTGCG CTGCGTANCC  
 401 GTCCGTTTCAG CAAAATCCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC  
 451 GTCAATGCGA TGGACACCAA TCCGCTNGCG GCAGACCTG TGTTGTGAT  
 30  
 501 CAAAGAAGCC GNCGANGATT TCAGACGANG TNTGCTGGTA TTGAGCCGTT  
 551 TGACCGAGCG TAAATCCAT GTGTGAAGG CAGCTGGCGC AGACGTGCCG  
 601 TCTGAAATG CTGCCAACAT CGAAACACAT GAATTCCGCG GCCCGCATCC  
 651 GGCGGTTTG AGTGGCACGC ACATTCATTT CATTGAGCCG GTCGGTGCAA  
 701 ACAAACCGT TTGGACCATC AATTATCAAG ATGTAATTGC CATCGGACGT  
 35  
 751 TTGTTTCAA CAGGCCGTCT GAACACCGAG CGCGTGATTG CTTTGGGTGG  
 801 TTCTCAAGTC AACAAACAC GCCTCTTGCG TACCGTTTG GGTGCGAAAT  
 851 TATCGCAAAT TACTGCGGGC GAATTGGTTG ACGCAGACAA CCGCGTGATT  
 901 TCCGGTTTCG TATTGAACGG CGCGATTACA CAAGGCGCGC ACGATTATTT  
 951 GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG  
 40  
 1001 AGCTGTTTCGG CTGGGTTGCG CCGCAGCCGG ACAAATACTC CATCACGCGT  
 1051 ACGACCCTCG GCCATTTCCT GAAAAACAAA CTCTTCAAGT TCACGACAGC  
 1101 CGTCAACGGT GGCAGCCGCG CCATGGTGCC GATTGGTACT TACGAGCGCG  
 1151 TAATGCCGCT AGACATCCTG CCTACCCTGC TTTTGC CGCA TTTAATCGTC  
 1201 GGCATACCG ACAGCGCGCA AGCATTGGGT TGCTTGAAT TGGACGAAGA  
 45  
 1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATANGGCC  
 1301 CGCTGTTGCG TAAGGTGCTG GAAACCNTTG AGAAGGAAGG CTGA

This encodes a protein having amino acid sequence <SEQ ID 128; ORF22a>:

50  
 1 MIKIKKGLNL PIAGRPEQVI YDGPVITEVA LLGEEYAGMR PSMKVKEGDA  
 51 VKKGQVLFED KXPGVVETA PVSCKIAAIH RGEKRVLSV VIAVEGNDEI  
 101 EFERYAPEAL ANLSGXEXXX NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF  
 151 VNAMDNTNPLA ADPVVIEKA XXDFRRXXLV LSRLTERKIH VCKAAGADVP  
 201 SENAAINIETH EFGGPHFAGL SGTHIHFIET VGANKTVWTI NYQDVIAIGR  
 251 LFATGRLNTE RVIALGGSQV NKPRLRLTVL GAKVSQITAG ELVDADNRVI  
 301 SGSVLNGAIT QGAHDYLGRI HNQISVIEEG RSKELFGWVA PQPDKYSITR  
 55  
 351 TTLGHFLKNK LFKFTAVNG GDRAMVPIGT YERVMPLDIL PTLLLRDLIV  
 401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EXGPLLRKVL ETXEKEG\*

The originally-identified partial strain B sequence (ORF22) shows 94.2% identity over a 158aa overlap with ORF22a:

60  
 orf22.pep MIKIKKGLNLPIAGRPEQAVYDGPATEVALLGEEYAGMRPSMKVKEGDAVKKQVLFED  
 orf22a MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMRPSMKVKEGDAVKKQVLFED

-124-

		10	20	30	40	50	60
		70	80	90	100	110	120
5	orf22.pep	KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYAPEALANLSGEEVRR					
	orf22a	KKXPGVVFTAPVSGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYAPEALANLSGXEXXX					
		70	80	90	100	110	120
10		130	140	150			
	orf22.pep	NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNP					
	orf22a	NLIQSGLWTALRXRPFSKIPAVDAEPFAIFVNAMDTNPLAADPVVVIKEAXXDFRRXXLV					
		130	140	150	160	170	180

The complete strain B sequence (ORF22-1) and ORF22a show 94.9% identity in 447 aa overlap:

15	orf22a.pep	10	20	30	40	50	60
	orf22-1	MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMRPXMVKVEGDAVKKGQVLFED      :     :     :     :     :     :     :     :					
20		10	20	30	40	50	60
	orf22a.pep	70	80	90	100	110	120
	orf22-1	KKXPGVVFTAPVSGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYAPEALANLSGXEXXX         :     :     :     :     :     :     :					
25		70	80	90	100	110	120
	orf22a.pep	130	140	150	160	170	180
	orf22-1	NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV      :     :     :     :     :     :     :					
30		130	140	150	160	170	180
	orf22a.pep	190	200	210	220	230	240
	orf22-1	LSRLTERKIHVCKAAGADVPSENAANIETHEFGGPHPAGLSGTHIHFI EPVGANKTVWTI      :     :     :     :     :     :     :					
35		190	200	210	220	230	240
	orf22a.pep	250	260	270	280	290	300
	orf22-1	NYQDVIAIGRLFATGRLINTERVIALGGSQVKNPRLRLTVLGAKVSQITAGELVDADNRVI      :     :     :     :     :     :     :					
40		250	260	270	280	290	300
	orf22a.pep	310	320	330	340	350	360
	orf22-1	SGSVLNGAITQGAHDYLG RYHNQISVIEEGRSKELFGWVAPQPKYSITRTTLGHFLKNK      :     :     :     :     :     :     :					
45		310	320	330	340	350	360
	orf22a.pep	370	380	390	400	410	420
	orf22-1	LFKFTTAVNGGDRAMVPIGTYERVMPLDILPTLLRLDLIVGDTDSAQALGCLELDEEDLA      :     :     :     :     :     :     :					
50		370	380	390	400	410	420
	orf22a.pep	430	440	LCSFVCPGKYEXGPLLRKVLETIEKEGX      :     :     :     :     :     :     :			
	orf22-1	LCSFVCPGKYEYGPLLRKVLETIEKEGX 430 440					
60		430	440				

Further work identified a partial gene sequence <SEQ ID 129> from *N.gonorrhoeae*, which encodes the following amino acid sequence <SEQ ID 130; ORF22ng>:

65	1	MIKIKKGLNL	PIAGRPEQVI	YDGPATEVA	LLGEEYVGM	PSMKIKEGEA
	51	VKKGQVLFED	KKNPGVVFTA	PASGKIAAIH	RGEKRVLSV	VI AVEGNDEI
	101	EFERYVPEAL	AKLSSEKVR	NLIQSGLWTA	LRTRPFSKIP	AVDAEPFAIF

151 VNAMDTNPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADVP  
 201 SENAANIETH EFGGPHAGL SGTHIHFI EP VGANKTVWTI NYQDVIAIGR  
 251 LFVTGRLNTE RVVALGGLQV NKPRLRLTVL GAKVSQLTAG ELVDADNRVI  
 301 SGSVLNGAIA QGAHDYLG RY HN\*

5 Further work identified complete gonococcal gene <SEQ ID 131>:

1 ATGATTAAAA TCAAAAAAGG TCTAAATCTG CCCATCGCGG GCAGACCGGA  
 51 GCAAGTCATT TATGACGGCC CGGCCATTAC CGAAGTCGCG TTGCTTGGCG  
 101 AAGAATATGT CGGCATGCGC CCCTCGATGA AAATCAAGGA AGGTGAAGCC  
 151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTAGT  
 201 ATTTACTGCG CCGGCTTCAG GCAAAATCGC CGCTATTAC CGTGGCGAAA  
 251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC  
 301 GAGTTCGAAC GCTACGTACC TGAAGCGCTG GCAAAATGTA GCAGCGAAAA  
 351 AGTGGCGCCG AACCTGATTC AATCAGGCTT ATGGACTGCG CTTGCGACCC  
 401 GTCCGTTTCA CAAAATCCCT GCCGTAGATG CCGAGCCGTT CGCCATCTTC  
 451 GTCAATGCGA TGGACACCAA TCCGCTGGCT GCCGACCCTA CGGTTCATCAT  
 501 CAAAGAAGCC GCCGAAGACT TCAAACGCGG CCTGTTGGTA TTGAGCCGCC  
 551 TGACCGAAGC TAAATCCAT GTGTGTAAAG CAGCAGGCGC AGACGTGCCG  
 601 CTGGAAGATG CTGCCAATAT CGAAACACAT GAATTTGGCG GCCCGCATCC  
 651 TGCCGGCTTG AGTGGCAGCG ACATTCATTT CATCGAGCCA GTCGGCGCGA  
 701 ATAAAACCGT GTGGACCATC AATTATCAAG ACGTGATTGC TATCGGACGT  
 751 TTGTTCTGTA CAGGCCGTCT GAATACCGAG CGCGTGGTTG CTTTGGGCGG  
 801 CCTGCAAGTC AACAAACGCG GCCTCTTGCG TACCGTTTGG GGTGCGAAGG  
 851 TGCTCAACT TACCGCCGCG GAATTGGTTG ACGCGGACAA CCGCGTGATT  
 901 TCCGTTTCGG TATTGAACGG TCGGATTGCA CAAGGCGCGC ATGATTATTT  
 951 GGGACGCTAC CACAATCAGA TTCCGTTTAT CGAAGAAGGC CGCAGCAAAG  
 1001 AGTGTTTCGG CTGGGTTGCG CCGCAGCCGG ACAAACTACTC CATCAGCGCG  
 1051 ACCACTCTCG GCCATTTCCT AAAAAACAAA CTCTTCAAGT TCACGACAGC  
 1101 CGTCAACGGC GCGCACCAGC CCATGGTACC GATCGGCACT TATGAGCGCG  
 1151 TAATGCCGTT GGACATCCTG CCTACCTTGC TTTTGCCGGA TTTAATCGTC  
 1201 GCGGATACCG ACAGCGCGCA GGCTTTGGGT TGCTTGAAT TGGACGAAGA  
 1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCG GGGCAAATAC GAATACGGCC  
 1301 CGCTGTTGCG CAAAGTGCTG GAAACCATTG AGAAGGAAGG CTGA

This encodes a protein having amino acid sequence <SEQ ID 132; ORF22ng-1>:

35 1 MIKIKKGLNL PIAGRPEQVI YDGAITEVA LLGEEYVGM RPSMKIKEGEA  
 51 VKKGQVLFED KKNPGVVETA PASGKIAAIH RGEKRVLSV VIAVEGNDEI  
 101 EFERYVPEAL AKLSSEKVR NLIQSLWTA LRTRPFSKIP AVDAEPFAIF  
 151 VNAMDTNPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADVP  
 201 SENAANIETH EFGGPHAGL SGTHIHFI EP VGANKTVWTI NYQDVIAIGR  
 40 251 LFVTGRLNTE RVVALGGLQV NKPRLRLTVL GAKVSQLTAG ELVDADNRVI  
 301 SGSVLNGAIA QGAHDYLG RY HNQISVIEEG RSKELEFWVA POPDKYSITR  
 351 TTLGHFLKNK LFKFTTAVNG GDRAMVPIGT YERVMLDIL PTLLLRDLIV  
 401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EYGPLLRLKVL ETIEKEG\*

The originally-identified partial strain B sequence (ORF22) shows 93.7% identity over a 158aa  
 45 overlap with ORF22ng:

orf22.pep MIKIKKGLNLPIAGRPEQAVYDGAITEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED 60  
 orf22ng MIKIKKGLNLPIAGRPEQVIYDGAITEVALLGEEYVGM RPSMKIKEGEAVKKGQVLFED 60  
 50 orf22.pep KKNPGVVETAPASGKIAAIHRGEKRVLSVVIAXNDEIEFERYAPEALANLSGEEVRR 120  
 orf22ng KKNPGVVETAPASGKIAAIHRGEKRVLSVVIAXNDEIEFERYVPEALAKLSSEKVR 120  
 55 orf22.pep NLIQSLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNP 158  
 orf22ng NLIQSLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV 180

The complete sequences from strain B (ORF22-1) and gonococcus (ORF22ng) show 96.2%  
 identity in 447 aa overlap:

60 orf22-1.pep 10 20 30 40 50 60  
 MIKIKKGLNLPIAGRPEQAVYDGAITEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED

-126-

	orff22ng-1	MIKIKKGLNLP 10	IAGRPEQVIYD 20	GPAITEVALLG 30	EELYVGM 40	PSMKIKEGE 50	AVKKGQVLF 60	FED
5	orff22-1.pep	70	80	90	100	110	120	
	orff22ng-1	70	80	90	100	110	120	
10	orff22-1.pep	130	140	150	160	170	180	
	orff22ng-1	130	140	150	160	170	180	
15	orff22-1.pep	190	200	210	220	230	240	
	orff22ng-1	190	200	210	220	230	240	
20	orff22-1.pep	250	260	270	280	290	300	
	orff22ng-1	250	260	270	280	290	300	
25	orff22-1.pep	310	320	330	340	350	360	
	orff22ng-1	310	320	330	340	350	360	
30	orff22-1.pep	370	380	390	400	410	420	
	orff22ng-1	370	380	390	400	410	420	
35	orff22-1.pep	430	440					
	orff22ng-1	430	440					

Computer analysis of these sequences gave the following results:

Homology with 48kDa outer membrane protein of *Actinobacillus pleuropneumoniae* (accession number U24492).

ORF22 and this 48kDa protein show 72% aa identity in 158aa overlap:

50	Orf22	1	MIKIKKGLNLP MI IKKGL+LPIAG P Q +++G + EVA+LGEEY GMRPSMKV+EGD VKKGQVLFED	60
	48kDa	1	MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGMPSMKVREGDVVKKGQVLFED	60
	orff22	61	KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI KKNPGVVFTAPASG + I+RGEKRVLSVVI VE +++I F RY LA+LS E+V++	120
55	48kDa	61	KKNPGVVFTAPASGTVVTINRGEKRVLSVVIKVEGDEQITFTRYEAAQLASLSAEQVKQ	120
	orff22	121	NLIQSGLWTALRTRPFSSKIPAVDAEPFAIFVNAMDTNP	158
	48kDa	121	NLI+SGLWTA RTRPFSSK+PA+DA P +IFVNAMDTNP	158
60			NLIESGLWTAFRTRPFSSKVPALDAIPSSIFVNAMDTNP	158

ORF22a also shows homology to the 48kDa *Actinobacillus pleuropneumoniae* protein:

gi|1185395 (U24492) 48 kDa outer membrane protein [Actinobacillus pleuropneumoniae]  
Length = 449

65 Score = 530 bits (1351), Expect = e-150

-127-

Identities = 274/450 (60%), Positives = 323/450 (70%), Gaps = 4/450 (0%)

Query: 1 MIKIKKGLNLPPIAGRPEQVIYDGPVITEVALLGEEYAGMRPXMKVKEGDAVKKGQVLFED 60  
 5 Sbjct: 1 MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGM RPSMKVREGDVVKKGQVLFED 60

Query: 61 KKXPGVVFTAPVSGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYAPEALANLSGXEXXX 120  
 10 Sbjct: 61 KKNPGVVFTAPASGT VVTINRGEKRVLSVVIKVEGDEQITFTRYEAAQLASLSAEQVKQ 120

Query: 121 NLIQSGLWTALRXRPFSKIPAVDAEPFAIFVNAMDTNPLAADPVVVIKEAXXDFRRXXLV 180  
 15 Sbjct: 121 NLIESGLWTAFTRPFSKVPALDAIPSSIFVNAMDTNPLAADPEVVLKEYETDFKDGLTV 180

Query: 181 LSRL--TERKIHVCKAAGADVP-SENAANIETHEFGGPHAGLSGTHIHFIIEPVGANKTV 237  
 20 Sbjct: 181 LTRLFNGQKPVYLCKDADSNIP LSPAIEGITIKSFSGVHPAGLVGTHIHFDVPVGATKQV 240

Query: 238 WTINYQDVIAIGRLFATGRLINTERVIALGGSQVKNPRLRLTVLGAKVSQITAGELVDADN 297  
 25 Sbjct: 241 WHLNYQDVIAIGKLETTGELFTDRIISLAGPQVKNPRLVRTRLGANLSQLTANELNAGEN 300

Query: 298 RVISGSVLNGAITQGAHDYLG RYHNQISVIEEGRSKELFGWVAPQDPKYSITRTTLGHFL 357  
 30 Sbjct: 301 RVISGSVLGATAAGPVDYLG RYALQVSVLAEGREKELFGWIMPGSDKFSITRTVLGHFG 360

Query: 358 KNKLFKFTTAVNGGDRAMVPIGT YERVMXXXXXXXXXXXXXXXXXVGDTDSAQXXXXXXXXXX 417  
 35 Sbjct: 361 K-KLFNFTTAVHGGERAMVPIGAYERVMPLDIIPTLLRLDLAAGDTDSAQNLGCLELDEE 419

Query: 418 XXXXS FVCPGKYEXGPLL RRVLETXEKEG 447  
 ++VCPGK GP+LR LE EKEG

ORF22ng-1 also shows homology with the OMP from *A.pleuropneumoniae*:

gi|1185395 (U24492) 48 kDa outer membrane protein [Actinobacillus  
 35 pleuropneumoniae] Length = 449  
 Score = 555 bits (1414), Expect = e-157  
 Identities = 284/450 (63%), Positives = 337/450 (74%), Gaps = 4/450 (0%)

Query: 27 MIKIKKGLNLPPIAGRPEQVIYDGP AITEVALLGEEYVGM RPSMKIKEGEAVKKGQVLFED 86  
 40 Sbjct: 1 MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGM RPSMKVREGDVVKKGQVLFED 60

Query: 87 KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYVPEALAKLSSEKVR 146  
 45 Sbjct: 61 KKNPGVVFTAPASG + I+RGEKRVLSVVI VEG+++I F RY LA LS+E+V++ 120

Query: 147 NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV 206  
 50 Sbjct: 121 NLIESGLWTAFTRPFSKVPALDAIPSSIFVNAMDTNPLAADPEVVLKEYETDFKDGLTV 180

Query: 207 LSRL--TERKIHVCKAAGADVP-SENAANIETHEFGGPHAGLSGTHIHFIIEPVGANKTV 263  
 55 Sbjct: 181 LTRLFNGQKPVYLCKDADSNIP LSPAIEGITIKSFSGVHPAGLVGTHIHFDVPVGATKQV 240

Query: 264 WTINYQDVIAIGRLFVTGRLINTERVIALGGLQVKNPRLRLTVLGAKVSQITAGELVDADN 323  
 60 Sbjct: 241 WHLNYQDVIAIGKLETTGELFTDRIISLAGPQVKNPRLVRTRLGANLSQLTANELNAGEN 300

Query: 324 RVISGSVLNGAIAQGAHDYLG RYHNQISVIEEGRSKELFGWVAPQDPKYSITRTTLGHFL 383  
 65 Sbjct: 301 RVISGSVLGATAAGPVDYLG RYALQVSVLAEGREKELFGWIMPGSDKFSITRTVLGHFG 360

Query: 384 KNKLFKFTTAVNGGDRAMVPIGT YERVMXXXXXXXXXXXXXXXXXVGDTDSAQXXXXXXXXXX 443  
 70 Sbjct: 361 K-KLFNFTTAVHGGERAMVPIGAYERVMPLDIIPTLLRLDLAAGDTDSAQNLGCLELDEE 419

Query: 444 XXXXS FVCPGKYEGPLL RRVLETIEKEG 473  
 ++VCPGK YGP+LR LE IEKEG

Sbjct: 420 DLALCTYVCPGKNYGPMLRAALEKIEKEG 449

Based on this analysis, including the homology with the outer membrane protein of *Actinobacillus pleuropneumoniae*, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF22-1 (35.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 5A shows the results of affinity purification of the GST-fusion protein, and Figure 5B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 5C). These experiments confirm that ORF22-1 is a surface-exposed protein, and that it is a useful immunogen.

## 10 Example 16

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 133>:

```

1  ..GCGnCGnAAA TCATCCATCC CC..nACGTC GTAGGCCCTG AAGCCAACTG
51  GTTTTTTATG GTAGCCAGTA CGTTTGTGAT TGCTTTGATT GGTTATTTTG
15  TTA CTGAAAA AATCGTCGAA CCGCAATTGG GCCCTTATCA ATCAGATTTG
151 TCACAAGAAG AAAAAGACAT TCGGCATTCC AATGAAATCA CGCCTTTGGA
201 ATATAAAGGA TTAATTTGGG CTGGCGTGGT GTTTGTTGCC TTATCCGCCC
251 TATTGGCTTG GAGCATCGTC CCTGCCGACG GTATTTTGCG TCATCCTGAA
301 ACAGGATTGG TTTCCGTTTC GCCGTTTTTA AAATCGATTG TTGTTTTTAT
351 TTTCTTGTTG TTTGCACTGC CGGGCATTGT TTATGGCCGG GTAACCCGAA
20 401 GTTTGCGCGG CGAACAGGAA GTCGTTAATG CGmyGGCCGA ATCGATGAGT
451 ACTCTGGsGC TTTmTTTGsw CAkCATCTTT TTTGCCGCAC AGTTTGTGCG
501 ATTTTTTAAT TGGACGAATA TTGGGCAATA TATTGCCGTT AAAGGGGCGA
551 CGTTCTTAAA AGAAGTCGGC TTGGGCGGCA GCGTGTGTGTT TATCGGTTTT
25 601 ATTTTAATTT GTGCTTTTAT CAATCTGATG ATAGGCTCCG CCTCCGCGCA
651 ATGGGCGGTA ACTGCGCCGA TTTTCGTCCC TATGCTGATG TTGGCCGGCT
701 ACGCGCCCGA AGTCATTCAA GCCGCTTACC GCATCGGTGA TTCCGTTACC
751 AATATTATTA CGCGGATGAT GAGTTATTTC GGGCTGATTA TGGCGACGGT
801 GrkCmmnTAC AAAAAGATG CGGGCGTGGG TaCGcTGATT wCTATGATG
851 TGCCGTATTC CGCTTTCTTC TTGATTGCgT GGATTGCCTT ATTCTGCATT
30 901 TGGGTATTTg TTTTGGGCCT GCCCGTCGGT CCCGGCGCGC CCACATTCTA
951 TCCCGCACCT TAA

```

This corresponds to the amino acid sequence <SEQ ID 134; ORF12>:

```

1  ..AXXIIHPXXV VGPEANWFFM VASTFVIALI GYFVTEKIVE PQLGPYQSDL
35 51  SQEEKDIRHS NEITPLEYKG LIWAGVVFVA LSALLAWSIV PADGILRHPE
101 TGLVSGSPFL KSIVVFIFLL FALPGIVYGR VTRSIRGEQE VVNAXAESMS
151 TLXLXLXXIF FAAQFVAFFN WTNIGQYIAV KGATFLKEVG LGGSVLFIGF
201 ILICAFINLM IGSASAQWAV TAPIFVPLM LAGYAPEVIQ AAYRIGDSVT
251 NIITPMMSYF GLIMATVXXY KKDAGVGTLI XMMLPYSAFF LIAWIALEFCI
301 WVFVLGLPVG PGAPTFFPAP *

```

40 Further sequence analysis revealed the complete DNA sequence <SEQ ID 135> to be:

```

1  ATGAGTCAAA CCGATACGCA ACGGGACGGA CGATTTTAC GCACAGTCGA
51  ATGGCTGGGC AATATGTTGC CGCATCCGGT TACGCTTTT ATTATTTCA
101 TTGTGTTATT GCTGATTGCC TCTGCCGTCG GTGCGTATTT CGGACTATCC
45 151 GTCCCCGATC CGGCCCTGT TGGTGCGAAA GGACGTGCCG ATGACGGTTT
201 GATTACATT GTCAGCCTGC TCAATGCCGA CGGTTTATC AAAATCCTGA
251 CGCATACCGT TAAAAATTC ACCGGTTTCG CGCCGTTGGG AACGGTGTG
301 GTTTCTTTAT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTTCCGC
351 ATTAATGCGC TTATTGCTCA CAAAATCGCC ACGCAAATC ACTACTTTTA
401 TGGTTGTTT TACAGGGATT TTATCTAATA CCGCTTCTGA ATTGGGCTAT
50 451 GTCGTCCTAA TCCCTTTGTC CGCCATCATC TTTCATTCCC TCGGCCGCCA
501 TCCGCTTGCC GGTCTGGCTG CGGCTTTCGC CGGCGTTTCG GCGGTTATT

```



**This corresponds to the amino acid sequence <SEQ ID 136; ORF12-1>:**

25

30

1	MSQDTQRDG	RFLRTVEWLG	NMLPHPVTLF	IIFIVLLLIA	SAVGAYFGLS
51	VPDRPVGAK	GRADDGLIYI	VSLLNADGFI	KILHTVKNF	TGFAPLGTVI
101	VSLLGVGIAE	KSGLSALMR	LLLTSPRKL	TFMVVFTGI	LSNTASELGY
151	VVLIPLSAII	FHSLGRHPLA	GLAAAFAGVS	GGYSANFLG	TIDPLLAGIT
201	QQAQIIHPD	YVVGPEANWF	EMVASTFVIA	LISYFVTEKI	VEPQLGPYQS
251	DLSEQEKIDR	HSNEITPLEY	KGLIWAGVVF	VALSALLAWS	IVPADGILRH
301	PETGLVSGSP	FLKSIVVFIF	LLFALPGIVY	GRVTRSLRGE	QEVVNAMAES
351	MSTLGLYLV	IFFAAQFVAF	FNWTNIGQYI	AVKGATFLKE	VGLGGSVLFT
401	GFILICAFIN	LMIGSASAQW	AVTAPIFVPM	LMLAGYAPEV	IQAAAYRIGDS
451	VTNIITPMMS	YFGLIMATVI	KYKKDAGVGT	LISMMLPEYS	FFLIAWIALE
501	CIWFEVLGLP	VGPAGATFYP	AP*		

Computer analysis of this amino acid sequence gave the following results:

35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF12 shows 96.3% identity over a 320aa overlap with an ORF (ORF12a) from strain A of *N. meningitidis*:

40  
45  
50  
55  
60

```

                                10           20           30
orf12.pep                      AXXI IHPXXVVGPEANWFFMVASTFVIALI
                                | | | | | | | | | | | | | | | | | |
orf12a      AAFAAGVSGGYSANLFLGTIDPLLAGITQAAQIIHPDYVVGPEANWFFMVASTFVIALI
              180           190           200           210           220           230

              40           50           60           70           80           90
orf12.pep      GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV
              | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf12a      GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV
              240           250           260           270           280           290

              100          110          120          130          140          150
orf12.pep      PADGILRHPETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAXAESMS
              | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf12a      PADGILRHPETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAMAESMS
              300          310          320          330          340          350

              160          170          180          190          200          210
orf12.pep      TLXLXLXXIFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIFIGFILICAFINLM
              | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf12a      TLGLYLVIIFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIFIGFILICAFINLM
              360          370          380          390          400          410

              220          230          240          250          260          270
orf12.pep      IGSASAQWAVTAPIFVPMMLLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVXXY

```

-130-

```

orf12a      IGSASQWAVTAPIFVPMMLLAGYAPEVIQAAAYRIGDSVTNIITPMMSYFGLIMATVIKY
              420      430      440      450      460      470

              280      290      300      310      320
5  orf12.pep  KKDAGVGTLLIXMMLPYSAFFLIWIALFCIWVFLVGLPVGPGAPTFFYPAPX
              |||||
orf12a      KKDAGVGTLLSMMLPYSAFFLIWIALFCIWVFLVGLPVGPGAPTFFYPAPX
              480      490      500      510      520

```

The complete length ORF12a nucleotide sequence <SEQ ID 137> is:

```

10      1  ATGAGTCAAA CCGATACGCA ACGGGACGGA CGATTTTTCAC GCACAGTCGA
      51  ATGGCTGGGC AATATGTTGC CGCACCCTGT TACGCTTTTTC ATTATTTTCA
     101  TTGTGTTATT GCTGATTGCC TCTGCCGCCG GTGCGTATTT CGGACTATCC
     151  GTCCCCGATC CGCGCCCTGT TGGTGCGAAA GGACGTGCCG ATGACGGTTT
     201  GATTACAGTT GTCAGCCTGC TCGATGCTGA CGGTTTGATC AAAATCCTGA
     15      251  CGCATACCGT TAAAAATTTT ACCGGTTTCG CGCCGTTGGG AACGGTGTG
     301  GTTTCCTTAT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTTCGCG
     351  ATTAATGCGC TTATTGCTCA CAAATCTCC ACGCAAATC ACTACTTTTA
     401  TGGTTGTTTT TACAGGATTT TTATCTAATA CCGCTTCTGA ATTGGGCTAT
     451  GTCGTCCTAA TCCCTTTGTC CGCCATCATC TTTCATTCCC TCGGCCGCCA
     20      501  TCCGCTTGCC GGTCTGGCTG CGGCTTTCGC CGGCGTTTCG GCGGTTATT
     551  CCGCCAATCT GTTCTTAGGC ACAATCGATC CGCTCTTGGC AGGCATCACC
     601  CAACAGGCGG CGCAAATCAT CCATCCCGAC TACGTCGTAG GCCCTGAAGC
     651  CCACTGGTTT TTTATGGTAG CCAGTACGTT TGTGATTGCT TTGATTGGTT
     701  ATTTTGTAC TGAAAAATC GTCGAACCGC AATTGGGCCC TTATCAATCA
     25      751  GATTGTGCAC AAGAAGAAAA AGACATTCGA CATTCCAATG AAATCAGGCC
     801  TTTGGAATAT AAAGGATTAA TTTGGGCTGG CGTGGTGTTC GTTGCCTTAT
     851  CCGCCCTATT GGCTTGGAGC ATCGTCCCTG CCGACGGTAT TTTGCGTCAT
     901  CCTGAAACAG GATTGGTTTC CGGTCGCCCG TTTTAAAT CAATTGTTGT
     951  TTTTATTTTC TTGTTGTTTG CACTGCCGGG CATTGTTTAT GGCCGGGTAA
     30      1001  CCCGAAGTTT GCGCGGCGAA CAGGAAGTCG TTAATGCGAT GGCCGAATCG
     1051  ATGACTACTC TGGGGCTTTA TTTGGTTCATC ATCTTTTTTG CCGCACAGTT
     1101  TGTCGCATTT TTAAATGGA CGAATATTGG GCAATATATT GCCGTTAAAG
     1151  GGGCGACGTT CTTAAAAGAA GTCGGCTTGG GCGGCAGCGT GTTGTATTATC
     1201  GGTTTTATTT TAATTTGTGC TTTTATCAAT CTGATGATAG GCTCCGCCTC
     35      1251  CGCGCAATGG GCGGTAATCG CGCCGATTTT CGTCCCTATG CTGATGTTGG
     1301  CCGGCTACGC GCCCGAAGTC ATTCAAGCCG CTTACCGCAT CGGTGATTCC
     1351  GTTACCAATA TTATTACGCC GATGATGAGT TATTTGCGGC TGATTATGGC
     1401  GACGGTGATC AAATACAAAA AAGATGCGGG CGTGGGTACG CTGATTCTTA
     1451  TGATGTTGCC GTATTCCGCT TTCTTCTTGA TTGCGTGGAT TGCCTTATTC
     40      1501  TGCATTGGG TATTTGTTT GGGCCTGCCC GTCGGTCCCG GCGCGCCAC
     1551  ATTCTATCCC GCACCTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 138>:

```

      1  MSQTDTRDGR FLRTVEWLG NMLPHVPTLF IIFIVLLLIA SAAGAYFGLS
     51  VPDPRPVGAK GRADDGLIHV VSLLDADGLI KILHTVKNF TGFAPLGTVL
     45      101  VSLLGVGIAE KSLGISALMR LLLTKSPERK TTFMVVFTGI LSNTASELGY
     151  VVLIPLSAII FHSLGRHPLA GLAAAFAGVS GGSYANLFLG TIDPLLAGIT
     201  QQAQIIHPD YVVGPEANWF FMVASTFVIA LIGYFVTEKI VEPQLGPYQS
     251  DLSQEEKDIR HSNEITPLEY KGLIWAGVVF VALSALLAWS IVPADGILRH
     301  PETGLVSGSP FLKSIVVFIF LLFALPGIVY GRVTRSLRGE QEVVNAMAES
     50      351  MSTLGLYLVI IFFAAQFVAF FNWTNIGQYI AVKGATFLKE VGLGGSVLEFI
     401  GFILICAFIN LMIGSASAQW AVTAPIFVPM LMLAGYAPEV IQAAYRIGDS
     451  VTNIITPMMS YFGLIMATVI KYKKDAGVGT LISMMLPYSA FFLIAWIALF
     501  CIWVFLVGLP VGPGAPTFFP AP*

```

55 ORF12a and ORF12-1 show 99.0% identity in 522 aa overlap:

```

      10      20      30      40      50      60
orf12a.pep  MSQTDTRDGRFLRTVEWLG NMLPHVPTLFI IIFIVLLLIASAAGAYFGLSVPDPRPVGAK
              |||||
orf12-1      MSQTDTRDGRFLRTVEWLG NMLPHVPTLFI IIFIVLLLIASAVGAYFGLSVPDPRPVGAK
              10      20      30      40      50      60

      70      80      90      100     110     120
orf12a.pep  GRADDGLIHVVSLLDADGLIKILHTVKNFTGFAPLGTVLVSLLGVGIAEKSLGISALMR
              |||||
orf12-1      GRADDGLIYVSLLNADGFIKILHTVKNFTGFAPLGTVLVSLLGVGIAEKSLGISALMR

```

-131-

		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf12a.pep	LLLT	KSPR	KLTT	FMVV	FTGIL	SNTASE
	orf12-1	LLLT	KSPR	KLTT	FMVV	FTGIL	SNTASE
		130	140	150	160	170	180
		190	200	210	220	230	240
10	orf12a.pep	GGYS	ANLFL	GTID	PLLAG	ITQQA	AIHP
	orf12-1	GGYS	ANLFL	GTID	PLLAG	ITQQA	AIHP
		190	200	210	220	230	240
		250	260	270	280	290	300
15	orf12a.pep	VEPQ	LGPI	QSDLS	QEEK	DIRHS	NEIT
	orf12-1	VEPQ	LGPI	QSDLS	QEEK	DIRHS	NEIT
		250	260	270	280	290	300
		310	320	330	340	350	360
20	orf12a.pep	PETG	LVSG	SPFL	KSIV	VFI	FLFAL
	orf12-1	PETG	LVSG	SPFL	KSIV	VFI	FLFAL
		310	320	330	340	350	360
		370	380	390	400	410	420
25	orf12a.pep	IFFAA	QFVA	FFNWT	NIGQY	IAVKG	ATFL
	orf12-1	IFFAA	QFVA	FFNWT	NIGQY	IAVKG	ATFL
		370	380	390	400	410	420
		430	440	450	460	470	480
30	orf12a.pep	AVTA	PIFV	PMLM	LAGYA	PEVI	QAAY
	orf12-1	AVTA	PIFV	PMLM	LAGYA	PEVI	QAAY
		430	440	450	460	470	480
		490	500	510	520		
35	orf12a.pep	LISM	LPYS	AFFL	IAWIA	LFCIW	FVLGL
	orf12-1	LISM	LPYS	AFFL	IAWIA	LFCIW	FVLGL
		490	500	510	520		

45 Homology with a predicted ORF from *N.gonorrhoeae*

ORF12 shows 92.5% identity over a 320aa overlap with a predicted ORF (ORF12.ng) from *N. gonorrhoeae*:

	orf12.pep	AXXIIHPXXVVGPEANWFFMVASTFVIALI	30
50	orf12ng	AAAFAGVSGGYSANLFLGTIDPLLAGITQQAQIIHPDYVVGPEANWFFMAASTFVIALI	232
	orf12.pep	GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALLAWSIV	90
55	orf12ng	GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALLAWSIV	292
	orf12.pep	PADGILRHPETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAXAESMS	150
	orf12ng	PADGILRHPETGLVAGSPFLKSIVVFIFLLFALPGIVYGRITRSLRGEREVVNAXAESMS	352
60	orf12.pep	TLXLXIXIFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIFIGFILICAFINLM	210
	orf12ng	TLGLYLVIIFFAAQFVAFFNWTNIGQYIAVKGAVFLKKFRLGGSVLFIFIGFILICAFINLM	412
65	orf12.pep	IGSASQWAVTAPIFVPMMLLAGYAPEVIQAAYRIGDSVTNIIIPMMSYFGLIMATVXXY	270
	orf12ng	IGSASQWAVTAPIFVPMMLLAGNAPQVIQAAYRIGDSVTNIIIPMMSYFGLIMATVIKY	472

```

orf12.pep      KKDAGVGTlixmmlpysaffliawialfciwvfvlgpvgpaptfypap  320
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf12ng        KKDAGVGTliSmmlpysaffliawialfciwvfvlgpvgpTPTfypvp  522

```

The complete length ORF12ng nucleotide sequence <SEQ ID 139> is:

```

5      1  ATGAGTCAAA CCGACGCGCG TCGTAGCGGA CGATTTTAC GCACAGTCGA
      51  ATGGCTGGGC AATATGTTGC CGCACCCGGT TACGCTTTT ATTATTTCA
     101  TTGTGTTATT GCTGATTGcc tctgCCGTCG GTGCGTATT CGGACTATCC
     151  GTCCCCGATC CGCGTCCTGT TGGGGCGAAA GGACGTGCCG ATGACGGTTT
     201  GATTCACGTT GTCAGCCTGC TCGATGCCGA CGGTTTGATC AAAATCCTGA
     251  CGCATACCGT TAAAAATTTT ACCGGTTTCG CGCCGTTGGG AACGGTGTTG
     301  GTTTCCTTTT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTTCCGC
     351  ATTAATGCGC TTATTGCTCA CAAAATCCCC ACGCAAATC ACTACTTTTA
     401  TGGTTGTTTT TACAGGGATT TTATCCAATA CGGCTTCTGA ATTGGGCTAT
     451  TCGTCCTTAA TCCCTTTGTC CGCCGTCATC TTTCATTTCG TCGGCCGCCA
     501  TCCGCTTGCC GGTGTTGGCTG CGGCTTTCGC CGGCGTTTCG GGCGGTTATT
     551  CGGCCAATCT GTTCTTAGGC ACAATCGATC CGCTCTTGCC AGGCATCACC
     601  CAACAGGCGG CGCAAATCAT CCATCCCGAC TACGTCGTAG GCCCTGAAGC
     651  CAACTGGTTT TTTATGGCAG CCAGTACGTT TGTGATTGCT TTGATTGGTT
     701  ATTTTGTTAC TGAAAAATC GTCGAACCGC AATTGGGCCC TTATCAATCA
     751  GATTGTGCAC AAGAAGAAAA AGACATTCGG CATTCCAATG AAATCAGCCC
     801  TTTGGAATAT AAAGGATTAA TTTGGGCAGG CGTGGTGTTC GTTGCCTTAT
     851  CCGCCCTATT GGCTTGAGC ATCGTCCCTG CCGACGGTAT TTTGCGTCAT
     901  CCTGAAACAG GATTGGTTGC CGGTCGCCG TTTTAAAT CGATTGTTGT
     951  TTTTATTTTC TTGTTGTTTG CGCTGCCGGG CATGTTTAT GGCCGGATAA
    1001  CCCGAAGTTT GCGCGGCGAA CGGGAAGTCG TTAATGCGAT GGCCGAATCG
    1051  ATGAGTACTT TGGGACTTTA TTTGGTCATC ATCTTTTTTG CCGCACAGTT
    1101  TGTGCGATTT TTTAATTGGA CGAATATTGG GCAATATATT GCCGTTAAAG
    1151  GGGCGGTGTT CTTAAAAGAA GTCGGCTTGG GCGGCAGTGT GTTGTATTATC
    1201  GGTTTTATTT TAATTTGTGC TTTTATCAAT CTGATGATAG GCTCCGCCTC
    1251  CGCGCAATGG GCGGTAACG CGCCGATTTT CGTCCCTATG CTGATGTTGG
    1301  CCGGCTACGC GCCCGAAGTC ATTCAAGCCG CTTACCGCAT CGGTGATTCC
    1351  GTTACCAATA TTATTACGCC GATGATGAGT TATTTGCGGC TGATTATGGC
    1401  GACGGTAATC AAATACAAAA AAGATGCGGG CGTAGGCACG CTGATTTCTA
    1451  TGATGTTGCC GTATTCCGCT TTCTTCTTAA TTGCATGGAT CGCCTTATTC
    1501  TGCATTGGG TATTTGTTT GGGTCGCCC GTCGGTCCC GCACACCAC
    1551  ATCTATCCG GTGCCTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 140>:

```

      1  MSQTDARRSG RFLRTVEWLG NMLPHPVTLF IIFIVLLLIA SAVGAYFGLS
     51  VPDPRPVGAK GRADDGLIHV VSLLDADGLI KILHTVKNF TGFAPLGTVL
    101  VSLLGVIAG KSLISALMR LLLTKSPRKL TTFMVVFTGI LSNTASELGY
    151  VVLIPLSAVI FHSLSGRHPLA GLAAAFAGVS GGYSANLFLG TIDPLLAGIT
    201  QQAAQIHPD YVVGPEANWF FMAASTEVIA LIGYFVTEKI VEPQLGPYQS
    251  DLSQEKDIR HSNEITPLEY KGLIWAGVVF VALSALLAWS IVPADGILRH
    301  PETGLVAGSP FLKSIVVFIF LLFALPGIVY GRITRSLRGE REVVNAMAES
    351  MSTLGLYLVI IFFAAQFVAF FNTNIGQYI AVKGAVFLKK FRLGGSVLEI
    401  GFILICAFIN LMIGSASAQW AVTAPIFVPM LMLAGNAPQV IQAAYRIGDS
    451  VTNITFMMS YFGLIMATVI KYKKGAVGT LISMMLPYSA FFLIAWIALF
    501  CIWVFLGLP VPGTPTTFYP VP*

```

ORF12ng shows 97.1% identity in 522 aa overlap with ORF12-1:

```

50      10      20      30      40      50      60
orf12-1.pep  MSQTDQTDGRFLRTVEWLG NMLPHPVTLF IIFIVLLLIASAVGAYFGLSVDPDRPVGAK
           |||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf12ng      MSQTDARRSGRFLRTVEWLG NMLPHPVTLF IIFIVLLLIASAVGAYFGLSVDPDRPVGAK

55      10      20      30      40      50      60
orf12-1.pep  GRADDGLIYIVSLLNADGFIKILHTVKNF TGFAPLGTVLVSLLGVIAGKSLISALMR
           |||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf12ng      GRADDGLIHVVSLLDADGLIKILHTVKNF TGFAPLGTVLVSLLGVIAGKSLISALMR

60      70      80      90      100     110     120
orf12-1.pep  GRADDGLIYIVSLLNADGFIKILHTVKNF TGFAPLGTVLVSLLGVIAGKSLISALMR
           |||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf12ng      GRADDGLIHVVSLLDADGLIKILHTVKNF TGFAPLGTVLVSLLGVIAGKSLISALMR

65      130     140     150     160     170     180
orf12-1.pep  LLLTKSPRKL TTFMVVFTGILSNTASELGYVVLIPLSAIFHSLSGRHPLAGLAAAFAGVS
           |||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf12ng      LLLTKSPRKL TTFMVVFTGILSNTASELGYVVLIPLSAIFHSLSGRHPLAGLAAAFAGVS

```

		130	140	150	160	170	180
		190	200	210	220	230	240
5	orf12-1.pep	GGYSANLFLGTIDPLLAGITQQAQIIHPDYVVGPEANWFFMVA	STFVIALIGYFVTEKI				
	orf12ng	GGYSANLFLGTIDPLLAGITQQAQIIHPDYVVGPEANWFFMAA	STFVIALIGYFVTEKI				
		190	200	210	220	230	240
		250	260	270	280	290	300
10	orf12-1.pep	VEPQLGPGYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFV	ALSALLAWSIVPADGILRH				
	orf12ng	VEPQLGPGYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFV	ALSALLAWSIVPADGILRH				
		250	260	270	280	290	300
		310	320	330	340	350	360
15	orf12-1.pep	PETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQ	EVVNMAESMSTLGLYLVI				
	orf12ng	PETGLVAGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGE	EVVNMAESMSTLGLYLVI				
		310	320	330	340	350	360
		370	380	390	400	410	420
	orf12-1.pep	IFFAAQFVAFNWTNIGQYIAVKGATFLKEVGLGGSVLF	FIGFILICAFINLMIGSASAQW				
	orf12ng	IFFAAQFVAFNWTNIGQYIAVKGAVFLKEVGLGGSVLF	FIGFILICAFINLMIGSASAQW				
25		370	380	390	400	410	420
		430	440	450	460	470	480
	orf12-1.pep	AVTAPIFVPMMLLAGYAPEVIAQAYRIGDSVTNIITPMMS	YFGLIMATVIKYKKGADGVT				
	orf12ng	AVTAPIFVPMMLLAGYAPEVIAQAYRIGDSVTNIITPMMS	YFGLIMATVIKYKKGADGVT				
30		430	440	450	460	470	480
		490	500	510	520		
	orf12-1.pep	LISMMLPYSAFFLIAWIALFCIWVFLGLPVGPGAPTFFYP	PAPX				
	orf12ng	LISMMLPYSAFFLIAWIALFCIWVFLGLPVGPGTPTFFYP	VPX				
35		490	500	510	520		

In addition, ORF12ng shows significant homology with a hypothetical protein from *E. coli*:

40	sp P46133 YDAH ECOLI HYPOTHETICAL 55.1 KD PROTEIN IN OGT-DBPA INTERGENIC REGION >gi 1787597 (AE000231) hypothetical protein in ogt 5'region [Escherichia coli] Length = 510 Score = 329 bits (835), Expect = 2e-89 Identities = 178/507 (35%), Positives = 281/507 (55%), Gaps = 15/507 (2%)
45	Query: 8 RSGRFLRTVEWLG NMLPHPV TXXXXXXXXXXASAVGAYFGLSVDP RPRV GAKGRADDGL 67 +SG+ VE +GN +PHP +A+ + FG+S +P D Sbjct: 13 QSGKLYGWVERIGNKVPHPFLLEFIYLIIVLMVTTAILSAFGVSAKNP-----TDGTP 64
50	Query: 68 IHVVSLLDADGLIKILTHTVKNFTGFAPXXXXXXXXXXIAEKSGLISALMRLLLTKSP 127 + V +LL +GL L + +KNE+GFAP +AE+ GL+ ALM + + Sbjct: 65 VVVKNLLSVEGLHWFLPNVIKNFSGFAPLGAILALVLGAGLAERVGLPALMVKMASHVN 124
55	Query: 128 RKLTTFMVVF TGILSNTASELG YVLIPLSAVIFHSLGRHPLAGLAAAFAGVSGGYSANL 187 + ++MV+F S+ +S+ V++ P+ A+IF ++GRHP+AGL AA AGV G++ANL Sbjct: 125 ARYASYMVLEIAFFSHISSDAALVIMPPMGALIFLAVGRHPVAGLLAAIAGVGCGETANL 184
60	Query: 188 FLGTIDPLLAGITQQAQIIHPDYVVGPEANWFFMAASTFVIALIGYFVTEKIVEPOLGP 247 + T D LL+GI+ +AA +P V NW+FMA+S V+ ++G +T+KI+EP+LG Sbjct: 185 LIVTTDVL LSGISTEAAA FNPQMHVSVIDNWYFMASSVVVLTIVGGLITDKIIEPRLGQ 244
65	Query: 248 YQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRH PETGLVA 307 +Q + ++ + + S GL AGVV + A +A ++P +GILR P V Sbjct: 245 WQNSDEKLQTLTESQRF-----GLRIAGVVVSLLFIAAIALMVIPQNGILRDPINHVTM 298
70	Query: 308 GSPFLKSIVVFIFLLFALPGIVYGRITRSLRGEREVVNMAESMSTLGLYLXXXXXXXXXX 367 SPF+K IV I L F + + YG TR++R + ++ + M E M + ++ Sbjct: 299 PSPFIKGIVPLIILFFVVS LAYGIATRTRIRQADLPHLMIEPMKEMAGFIVMVFEPLAQF 358
	Query: 368 XXXXNWTNIGQYIAVKGAVFLKEVGLGGSVLFIFIGFILICAFINLMIGSASAQWAVTAPIF 427 NW+N+G++IAV L+ GL G F+G L+ +F+ + I S SA W++ APIF

-134-

Sbjct: 359 VAMFNWSNMGKFIAVGLTDILESSGLSGIPAFVGLALLSSFLCMFIASGSAIWSILAPIF 418  
 Query: 428 VPMLMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGTLISMMLP 487  
 VPM ML G+ P Q +RI DS + P+ + L + + +YK DA +GT S++LP  
 Sbjct: 419 VPMFMLLGfHPAFAQILFRIADSSVLPLAPVSPFVPLFLGFLQRYKPKDAKLGTYYSLVLP 478  
 Query: 488 YSAFFLIAWIALFCIWVFLGLPVGPG 514  
 Y FL+ W+ + W +++GLP+GPG  
 Sbjct: 479 YPLIFLVVWLLMLLAW-YLVGLPIGPG 504

Based on this analysis, including the presence of several putative transmembrane domains and the predicted actinin-type actin-binding domain signature (shown in bold) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 17

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 141>:

```

1  ..ACAGCCGGCG CAGCAGGTTn CnCGGTCTTC GTTTTCGTAA CGGACAGTCA
51  GGTGGAGGTG TTCGGAACA TCCAGACCGC AGTGGAAACA GGTTTTTTTC
101 ATGGCATTTC GGTTTCGTCT GTGTTTGGTG CGGCGGCACA AGACTCGGCA
151 ATgGCTTCGC GCAGTGCCTC TATACCGGTA TTTTCAGCAA CGGAAATGCG
201 GACGGcGgCA ATTTTCCCG CAGCGTCGCG CCATATGCCC GTGTTTgTT
251 CTTCAGACGG CAGCAGGTCG GTTTTGTGT ACACCTTgAT GCACGGaATA
301 TCGCCGGCAT GGATTTCTTG CAGTACGTTT TCCACGTCTT CAATCTGCTG
351 TCCGCTGTTT GGAGCGGCGG CATCGACGAC GTGCAGCAGC ACATCgGcTT
401 gCGCGGTTTC TTCCAGCGTG GCgGAAAAGG CGGAAATCAG TTTgTGCggc
451 agATyGCTnA CGAATCCGAC GGTATCGGTC AGGATAATGC TGCATTCGGG
501 ACT..

```

This corresponds to the amino acid sequence <SEQ ID 142; ORF14>:

```

1  ..TAGAAGXXVF VFVTD SQVEV FGNIQTAVET GFFHGISVSS VFGAAQDSA
51  MASRSASIPV FSATEMRTAA IFPAASRHMP VFCSSDGSRs VLLYTLMHGI
101 SPAWISCSTF STSSICPLF GAAASTTCSS TSACAVSSSV AEKAEISLCG
151 RXLTNP TVSV RIMLHSG..

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF14 shows 94.0% identity over a 167aa overlap with an ORF (ORF14a) from strain A of *N.meningitidis*:

				10	20	30		
orf14.pep				TAGAAGXXVFVFVTD	SQVEVFGNIQTAVET			
				:	:			
40	orf14a	GRQLGFLRVGGALFVITAQARVNNALCDCLTTGAAGFAVFVFDGQM	QVFGNVQPAVET					
		150	160	170	180	190	200	
			40	50	60	70	80	90
45	orf14.pep	GFFHGISVSSVFGAAQDSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRs						
	orf14a	GFFHGISVSSVFGAAQYSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRs						
		210	220	230	240	250	260	
			100	110	120	130	140	150
50	orf14.pep	VLLYTLMHGISP	AWISCSTFSTSSICPLFGAAASTTCSSTSACAVSSSV	AEKAEISLCG				
	orf14a	VLLYTLMHGISP	AWISCSTFSTSSICPLFGAAASTTCSSTSACAVSSSV	AEKAEISLCG				
		270	280	290	300	310	320	

		1	ATGGAGGATT	TGCAGGAAAT	CGGGTTCGAT	GTCGCCGCCG	TAAAGGTAGG
	51	TCGGCAGCGC	GAACATCATC	GTCTGCATCA	TCCCCAGCCC	GGCAACGGCG	
10	101	AGGCGGACGA	TGTATTGTTT	CGGTTCTTTT	TGGTTGGCGG	CTTCGATTTT	
	151	TTGCGCGTCA	TAGGTCGCG	CGGTGTAGCC	TATCTGCCTG	ATTTTCAACA	
	201	GAATGTCGGA	AAGGCGGATT	TTGCGGTCGT	CCCGACGACG	CGGGACGCGG	
	251	TGCGTGCTGT	AATTGAGGTC	GATGCGGACG	ATGCCGTCTG	TACGCAAAAG	
	301	CTGCTGTTCC	ATCAGCCAGA	CGCAGGCGGC	GCAGGTGATG	CCGCCGAGCA	
15	351	TTAAAACCGC	CTCGCGCTG	CCGCCGTGGG	TTTCCACAAA	CTCGGACTGG	
	401	ACTTCGGGCA	GGTCGTACAG	CGCGATTGGG	TCGAGGATTT	TTTGGGCGGG	
	451	CAGCTCGGTT	TTTTGCGCGT	CGGCGGTGCG	TTGTTTGTAA	TAAGTCGCCA	
	501	AGCCCGCGTC	AATAATGCTT	TGTGCGACTG	CCTGACAACC	GGCGCAGCAG	
	551	GTTTCGCGGT	CTTCGTTTTC	GTAACGGACG	GTCAGATGCA	GGTTTTCGGG	
20	601	AACGTCACG	CCGCAGTGGA	AACAGGTTTT	TTTCATGGCA	TTTCGGTTTC	
	651	GTCTGTGTTT	GGTGCGCGGG	CACAATACTC	GGCAATGGCT	TCGCGCAGTG	
	701	CGTCTATACC	GGTATTTTCA	GCAACGGAAA	TGCGGACGGC	GGCAATTTTT	
	751	CCCGCAGCGT	CGCGCCATAT	GCCCGTGTTT	TGTTCTTCAG	ACGGCAGCAG	
	801	GTCGGTTTTG	TTGTACACAT	TGATGCACGG	AATATCGCCG	GCATGGATTG	
	851	CTTGACAGTAC	TTTTTCCACG	TCTTCAATCT	GCTGTCGCGT	GTTCCGAGCG	
25	901	GCGGCATCGA	CGACGTGCAG	CAGCACATCG	GCTTGCGCGG	TTTCTTCCAG	
	951	CGTGGCGGAA	AAGGCGGAAA	TCAGTTTGTG	CGGCAGATCG	CTGACGAATC	
	1001	CGACGGTATC	GGTCAGGATA	ATGCTGCATT	CGGGACTTGT	GTACAGCCGC	
	1051	CGCGCCGTGC	TGTCGAGTGT	GGCGAAAAGC	TGGTCTTTAG	CATATATGCC	
	1101	CGACTTTGGTC	AGCCGGTTGA	ACAGACTTGA	TTTGGCGACA	TTGGTATAG	

	1	MEDLQEIGFD	VAAVKVGRQR	EHHRLHHPQP	GNGEADDVLF	AFFLVGGFDF
	51	LRVIGCGGVA	YLPDFQQNVG	KADFAVVPDD	AAAVRAVIEV	DADDAVCTQK
	101	LLFDQPDAGG	AGDAAEH*NR	LARAAVGFHK	VGLDFQGVVQ	ADLVEDFLGR
	151	QLGFLRVVGA	LFVITAQARV	NNALCDCLT	GAGFAVVFV	VTDGQMQVFG
35	201	NVQPAVETGF	FHGISVSSVF	GAAAQYSAMA	SRSASIPVFS	ATEMRTAIIF
	251	PAASRHMPVF	CSSDGRSVL	LYTLMHGISP	AWISCSTFST	SSICCPLEGA
	301	AASTTCSSTS	ACAVSSSVAE	KAEISLCGRS	LTNPVTSVRI	MLHSGLMYSR
	351	RAVSSSVAKS	WSEAYMPDLV	SRINRLDLPT	LV*	

40 Homology with a predicted ORF from *N.gonorrhoeae*

	orf14.pep		TAGAAGXXVFEVFTDSQVEVFGNIQTAVET		30
			: : : : : : :		
45	orf14.ng	GRQFGFFRVGGASFEVITAQAGIDDALCDCLTADAAGFAVFAFVADGQMVFQGNVQPAVET			208
	orf14.pep	GFFHGISVSSVFGAAAQDSAMASRSASIPVFSATEMRTAAIFPAASRHPVFCSSDGSRS			90
50	orf14.ng	GFFHGISVSSVFGAAQYSAMASRSASIPVFSATEMRTAAIFPAASRHPVFCSSDGSRS			268
	orf14.pep	VLLYTLMHGISPAWISCSFTSTSSICPLFGAAASTTCSSTSACAVSSSVAEKAIEISLCG			150
	orf14.ng	VLLYTLMHGISWAWISCSFTSTSSICPLFRAAASTTCSSTSACTVSSKVAEKAIEISLCG			328
55	orf14.pep	RXLNTPTVSVRIMLHSG			167
	orf14.ng	RLNTPTVSVRIMLHAGLMYSRRVAVSRVAKSWSFAYMPDLVSRLNRLDLPTIV			382

The complete length ORF14ng nucleotide sequence <SEQ ID 145> is predicted to encode a protein having amino acid sequence <SEQ ID 146>:

1 MEDLQEIGFD VAAVKVGRQR EHHRLHHTQS GNGKADDVLF AFFLVGGFDF  
 51 LRVIGCGGVA CLPDFQONVG EADFAVVPDD AAARAVVIEV DADDAVCAQK  
 101 LLFDQPDAGG AGNAAEHQHC FVRAIMGFHK VGLDFGQVQVQ ADLVEDFLGR  
 151 QFGFFRVGGA SFVITAQAGI DDALCDCLTA DAAGFAVFAF VADGQMQRVFG  
 5 NVQPAVETGF FHGISVSSVF GAAQYSAMA SRSASIPVFS ATEMRTAAIF  
 201 PAASRHMPVF CSSDGSRSVL LYTLMHGISW AWISCSTFST SSICCPLEFRA  
 251 AASTTCSSTS ACTVSSKVAE KAEISLCGRS LTNPVSVRI MLHAGLMYSR  
 301 RAVVSRVAKS WSFAYMPDLV SRLNRLDLPT LV\*

Based on the putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 18

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 147>:

1 ..GGCCATTACT CCGACCGCAC TTGGAAGCCG CGTTTGGNCG GCCGCCGTCT  
 15 51 GCCGTATCTG CTTTATGGCA CGCTGATTGC GGTATTGTG ATGATTTGA  
 101 TGCCGAACATC GGGCAGCTTC GGTTTCGGCT ATGCGTCGCT GCGCGCTTG  
 151 TCGTTCCGGCG CGCTGATGAT TCGCGTGTGA GACGTGTCGT CAAATATGGC  
 201 GATGCAGCCG TTTAAGATGA TGGTCGGCGA CATGGTCAAC GAGGAGCAGA  
 251 AAA.NTACGC CTACGGGATT CAAAGTTTCT TAGCAAATAC GGGCGCGGTC  
 20 301 GTGGCGGCGA TTCTGCCGTT TGTGTTTGC TATATCGGTT TGGCGAACAC  
 351 CGCCGANAAA GCGCTTGTGC CGCAGACCGT GGTCGTGGCG TTTTATGTGG  
 401 GTGCGGCGTT GCTGGTGATT ACCAGCGCGT TCACGATTTT CAAAGTGAAG  
 451 GAATACGANC CGGAAACCTA CGCCCGTTAC CACGGCATCG ATGTCGCCCG  
 501 GAATCAGGAA AAAGCCAAC TGATCGCACT CTTAAAA.CC GCGC..

25 This corresponds to the amino acid sequence <SEQ ID 148; ORF16>:

1 ..GHYSDRTWKP RLXGRRLPYL LYGTLIAVIV MILMPNSGSF GFGYASLAL  
 51 SFGALMIALL DVSSNMAMQP FKMMVGDVNV EEQKXYAYGI QSFLANTGAV  
 101 VAAILPFVFA YIGLANTAXK GVVPQTVVVA FYVGAALLVI TSAFTIFKVK  
 151 EYXPETYARY HGIDVAANQE KANWIALLLKX A..

30 Further work revealed the complete nucleotide sequence <SEQ ID 149>:

1 ATGTCGGAAT ATACGCCTCA AACAGCAAAA CAAGGTTTGC CCGCGCTGGC  
 51 AAAAAGCACG ATTTGGATGC TCAGTTTCGG CTTTCTCGGC GTTCAGACGG  
 101 CCTTTACCTT GCAAAGCTCG CAAATGAGCC GCATTTTTC AAGCTAGGC  
 151 GCAGACCCGC ACAATTGGG CTGGTTTTC ATCTGCGCG CGCTGGCGGG  
 35 201 GATGCTGGTG CAGCCGATTG TCGGCCATTA CTCCGACCGC ACTTGAAGC  
 251 CGCGTTTGGG CGGCCGCCGT CTGCCGTATC TGCTTTATGG CACGCTGATT  
 301 GCGGTTATTG TGATGATTTT GATGCCGAAC TCGGGCAGCT TCGGTTTCGG  
 351 CTATGCGTCG CTGGCGGCTT TGTGTTTCGG CGCGCTGATG ATTGCGCTGT  
 401 TAGACGTGTC GTCAAATATG GCGATGCAGC CGTTTAAGAT GATGGTCGGC  
 40 451 GACATGGTCA ACGAGGAGCA GAAAGGCTAC GCCTACGGGA TTCAAAGTTT  
 501 CTTAGCAAAT ACGGGCGCGG TCGTGGCGGC GATTCTGCCG TTTGTGTTTG  
 551 CGTATATCGG TTTGGCGAAC ACCGCCGAGA AAGGCGTTGT GCCGCAGAC  
 601 GTGGTCGTGG CGTTTATGT GGGTGCGGCG TTGCTGGTGA TTACCAGCGC  
 651 GTTCACGATT TTCAAAGTGA AGGAATACGA TCCGGAACCC TACGCCGCTT  
 45 701 ACCACGGCAT CGATGTCGCC GCGAATCAGG AAAAAGCCAA CTGGATCGAA  
 751 CTCTTGA AAA CCGCGCCTAA GCGGTTTGG ACGGTTACTT TGGTGCAATT  
 801 CTTCTGCTGG TTCGCTTCC AATATATGTG GACTTACTCG GCAGGCGCGA  
 851 TTGCGGAAAA CGTCTGGCAC ACCACCGATG CGTCTTCCGT AGGTTATCAG  
 901 GAGGCGGGTA ACTGTTACGG CGTTTGGCG GCGGTGCACT CGGTTGCGGC  
 50 951 GGTGATTTGT TCGTTGTAT TGGCGAAAGT GCCGAATAAA TACCATAAGG  
 1001 CGGGTTATTT CGGCTGTTTG GCTTTGGGCG CGCTCGGCTT TTTCTCGGTT  
 1051 TTCTTCATCG GCAACCAATA CGCGCTGGTG TTGCTTATA CCTTAATCGG  
 1101 CATCGCTTGG GCGGGCATT TCACTTATCC GCTGACGATT GTGACCAACG  
 1151 CCTGTGCGGG CAAGCATATG GGCACCTACT TGGGCTTGT TAACGGCTCT  
 55 1201 ATCTGTATGC CTCAAATCGT CGCTTCGCTG TTGAGTTTCG TGCTTTTCCC  
 1251 TATGCTGGG GCGCTTCAGG CCACTATGTT CTTGGTAGGG GCGCTCGTCC  
 1301 TGCTGCTGGG CGCGTTTTC GTGTTCTCTGA TTAAAGAAAC ACACGGCGGG  
 1351 GTTTGA



This corresponds to the amino acid sequence <SEQ ID 150; ORF16-1>:

```

1  MSEYTPQTAK QGLPALAKST IWMLSFGLG VQTAFTLQSS QMSRIFQTLG
51  ADPHNLGWFF ILPPLAGMLV QPIVGHYSR TWKPRLGRR LPYLLYGTLI
101 AVIVMILMPN SGSFGFGYAS LAALSFGALM IALLDVSSNM AMQPFKMMVG
5  151  DMVNEEQGY AYGIQSFLAN TGAUVAAILP FVFAYIGLAN TAEKGVVPQT
201  VVVAFYVGAA LLVITSFTI FKVKEYDPET YARYHGIDVA ANQEKANWIE
251  LLKTAPKAFW TVTLVQFFCW FAFQYMWYTS AGAIAENVWH TTDASSVGYQ
301  EAGNWWGVLA AVQSVAVIC SFVLAKVPNK YHKAGYFGCL ALGALGFFSV
351  FFIGNQYALV LSYTLIGIAW AGIITYPLTI VTNALSGKHM GTYLGLENGS
10 401  ICMQPIVASL LSFVLFPMGL GLOATMFLVG GVVLLGAFS VFLIKETHGG
451  V*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF16 shows 96.7% identity over a 181aa overlap with an ORF (ORF16a) from strain A of *N.*

15 *meningitidis*:

```

                                10      20      30
orf16.pep                      GHYSDRTWKPRXLXGRRLLPYLLYGTLIAVIV
20 orf16a                      IFQTLGADPHSLGWFFILPPLAGMLVQPIVGHYSDRTWKPRLGRRLLPYLLYGTLIAVIV
                                50      60      70      80      90      100

                                40      50      60      70      80      90
orf16.pep                      MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDVNEEQKXYAYGI
25 orf16a                      MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDVNEEQKGYAYGI
                                110     120     130     140     150     160

                                100     110     120     130     140     150
orf16.pep                      QSFLANTGAVVAAILPFVFAYIGLANTAXKGVVPQTVVVAFYVGAALLVITSFTIFKVK
30 orf16a                      QSFLANTGAVVAAILPFVFAYIGLANTAEGVVPQTVVVAFYVGAALLVITSFTIFKVK
                                170     180     190     200     210     220

                                160     170     180
orf16.pep                      EYXPETYARYHGIDVAANQEKANWIALLKXA
35 orf16a                      EYNPETYARYHGIDVAANQEKANWIELLKTAPKAFWTVTLVQFFCWFAFQYMWYTSAGAI
                                230     240     250     260     270     280

40 orf16a                      AENVWHTTDASSVGYQEAGNWWGVLA AVQSVAVICSFVLAKVPNKYHKAGYFGCLALGA
                                290     300     310     320     330     340

```

The complete length ORF16a nucleotide sequence <SEQ ID 151> is:

```

1  ATGTCGGAAT ATACGCCTCA AACAGCAAAA CAAGGTTTGC CCGCGCTGGC
45 51  AAAAAGCACG ATTTGGATGC TCAGTTTCGG CTTTCTCGGC GTTCAGACGG
101 CCTTACCCT GCAAAGCTCG CAGATGAGCC GCATCTTCCA GACGCTCGGT
151 GCCGATCCGC ACAGCCTCGG CTGGTTCTTT ATCCTGCCGC CGCTGGCGGG
201 GATGCTGGTG CAGCCGATTG TCGGCCATTA CTCCGACCGC ACTTGAAGC
251 CGCGTTTGGG CGGCCGCGT CTGCCGTATC TGCTTTATGG CACGCTGATT
50 301 GCGGTTATTG TGATGATTTT GATGCCGAAC TCGGGCAGCT TCGGTTTCGG
351 CTATGCGTCG CTGGCGGCTT TGTGCTTCGG CGCGCTGATG ATTGCGCTGT
401 TAGACGTGTC GTCAAATATG GCGATGCAGC CGTTTAAGAT GATGGTCGGC
451 GACATGGTCA ACGAGGAGCA GAAAGGCTAC GCCTACGGGA TTCAAAGTTT
501 CTTAGCGAAT ACGGGCGCGG TCGTGCGGCG GATTCTGCCG TTTGTGTTTG
55 551 CGTATATCGG TTTGGCGAAC ACCGCCGAGA AAGGCGTTGT GCCGCAGACC
601 GTGGTCGTGG CGTTTATGTA GGGTGCGGCG TTGCTGTGTA TTACCAGCGC
651 GTTACGATT TTCAAAGTGA AGGAATACAA TCCGGAACCC TACGCCCGTT
701 ACCACGGCAT CGATGTCGCC GCGAATCAGG AAAAAGCCAA CTGGATCGAA
751 CTCTGAAAA CCGCGCCTAA GCGGTTTGG ACGGTTACTT TGGTGCAATT
801 CTTCTGCTGG TTCGCCTTCC AATATATGTG GACTTACTCG GCAGGCGCGA
60 851 TTGCGGAAAA CGTCTGGCAC ACCACCGATG CGTCTCCCGT AGGTTATCAG
901 GAGCGGGTA ACTGGTACGG CGTTTGGCG GCGGTGCAGT CGGTTGCGG
951 GGTGATTTGT TCGTTTGTAT TGGCGAAAGT GCCGAATAAA TACCATAAGG

```

-138-

5  
 1001 CGGGTTATTT CGGCTGTTTG GCTTTGGGCG CGCTCGGCTT TTTCTCCGTT  
 1051 TTCTTCATCG GCAACCAATA CGCGCTGGTG TTGTCTTATA CCTTAATCGG  
 1101 CATCGCTTGG GCGGGCATA TCACTTATCC GCTGACGATT GTGACCAACG  
 1151 CCTTGTCTGG CAAGCATATG GGCCTTACT TGGGCCTGTT TAACGGCTCT  
 1201 ATCTGTATGC CGCAAATCGT CGCTTCGCTG TTGAGTTTCG TGCTTTTCCC  
 1251 TATGCTGGGC GGCTTGCAGG CCACTATGTT CTTGGTAGGG GCGCTCGTCC  
 1301 TGCTGCTGGG CGCGTTTTCC GTGTTCTCTGA TTAAAGAAAC ACACGGCGGG  
 1351 GTTTGA

This encodes a protein having amino acid sequence <SEQ ID 152>:

10  
 1 MSEYTPQTAK QGLPALAKST IWMLSFGFLG VQTAFTLQSS QMSRIFQTLG  
 51 ADPHSLGWFF ILPLAGMLV QPIVGHYSR TWKPRLGRR LPYLLYGTLI  
 101 AVIVMILMPN SGSFGFGYAS LAALSFALM IALLDVSSNM AMQPFKMMVG  
 151 DMVNEEQGY AYGIQSFLAN TGAVVAAILP FVFAYIGLAN TAEKGVVPQT  
 201 VVVAFYVGAA LLVITSFTI FKVKYNPET YARYHGIDVA ANQEKANWIE  
 15  
 251 LLKTAPKAFW TVTLVQFFCW FAFQYMWYTS AGAIAENVWH TTDASSVGYQ  
 301 EAGNWWYGLA AVQSVAVIC SFVLAKVPNK YHKAGYFGCL ALGALGFFSV  
 351 FFIGNQYALV LSYTLIGIAW AGIITYPLTI VTNALSGKHM GTYLGLENFS  
 401 ICMPQIVASL LSFVLFPM LGQATMFLVG GVVLLGAFS VFLIKETHGG  
 451 V\*

20 ORF16a and ORF16-1 show 99.6% identity in 451 aa overlap:

		10	20	30	40	50	60
orf16a.pep		MSEYTPQTAKQGLPALAKSTIWMLSFGFLGVQTAFTLQSSQMSRIFQTLGADPHSLGWFF					
orf16-1		MSEYTPQTAKQGLPALAKSTIWMLSFGFLGVQTAFTLQSSQMSRIFQTLGADPHNLGWFF					
		10	20	30	40	50	60
25							
		70	80	90	100	110	120
orf16a.pep		ILPLAGMLVQPIVGHYSRDTWKPRLGRRLPYLLYGTLI					
orf16-1		ILPLAGMLVQPIVGHYSRDTWKPRLGRRLPYLLYGTLI					
		70	80	90	100	110	120
30							
		130	140	150	160	170	180
orf16a.pep		LAALSFALMIALLDVSSNMAMQPFKMMVGD					
orf16-1		LAALSFALMIALLDVSSNMAMQPFKMMVGD					
		130	140	150	160	170	180
35							
		190	200	210	220	230	240
orf16a.pep		FVFAYIGLANTA					
orf16-1		FVFAYIGLANTA					
		190	200	210	220	230	240
40							
		250	260	270	280	290	300
orf16a.pep		ANQEKANWIELLKTAPKAFWTVTLVQFFCWFAFQYMWYTSAGAIAENVWHTTDASSVGYQ					
orf16-1		ANQEKANWIELLKTAPKAFWTVTLVQFFCWFAFQYMWYTSAGAIAENVWHTTDASSVGYQ					
		250	260	270	280	290	300
45							
		310	320	330	340	350	360
orf16a.pep		EAGNWWYGLAAVQSVAVICSFVLAKVPNKYHKAGYFGCLALGALGFFSVFFIGNQYALV					
orf16-1		EAGNWWYGLAAVQSVAVICSFVLAKVPNKYHKAGYFGCLALGALGFFSVFFIGNQYALV					
		310	320	330	340	350	360
50							
		370	380	390	400	410	420
orf16a.pep		LSYTLIGIAWAGIITYPLTIVTNALSGKHMGTYLGLFNGSICMPQIVASLLSFVLFPM LG					
orf16-1		LSYTLIGIAWAGIITYPLTIVTNALSGKHMGTYLGLFNGSICMPQIVASLLSFVLFPM LG					
		370	380	390	400	410	420
55							
		430	440	450			
orf16a.pep		GLQATMFLVGGVVLLGAFSVFLIKETHGGVX					
orf16-1		GLQATMFLVGGVVLLGAFSVFLIKETHGGVX					
		430	440	450			
60							
65							

Homology with a predicted ORF from *N.gonorrhoeae*

ORF16 shows 93.9% identity over a 181aa overlap with a predicted ORF (ORF16.ng) from *N.gonorrhoeae*:

```

5      orf16.pep                      GHYSDRTWKPRXLXGRRLLPYLLYGTLIIVIV  30
      orf16ng      HFSNARRRPAQFGLVFHPAAAGGDAGSADSGYSDRTWKPRLGGRRLLPYLLYGTLIIVIV  131
10     orf16.pep      MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDMVNEEQKXYAYGI  90
      orf16ng      MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDMVNEEQKXYAYGI  191
      orf16.pep      QSFLANTGAVVAAILPFVFAYIGLANTAXKGVPVQTVVVAFYVGAALLVITSFTIFKVK  150
15     orf16ng      QSFLANTDAVVAAILPFVFAYIGLANTAEGVVPVQTVVVAFYVGAALLIITSFTISKVK  251
      orf16.pep      EYXPETYARYHGIDVAANQEKANWIALLKXA  181
      orf16ng      EYDPETYARYHGIDVAANQEKANWFELLKTAPKVFVTVTFVQFFCWFAFRYMWYTYSAGAI  311

```

20 The complete length ORF16ng nucleotide sequence <SEQ ID 153> is:

```

      1 ATGATAGGGG ATCGCCGCGC CGGCAACCAT TTCGGATTTT CCRAAGCAAA
      51 TACTTTTCAA ATCAAAAAAA AGGATTTACT TTATGTCGGA ATATACGCCT
     101 CAAACAGCAA AACAAGGTTT GCCCGCGCCG GCAAAAAGCA CGATTGGAT
     151 GTTGAGCTTC GGCTATCTCG GCGTTCAGAC GGCCTTTACC CTGCAAAGCT
     201 CGCAGATGAG CCGCATTTTT CAAACGCTAG GCGCAGACCC GCACAATTTG
     251 GGCTGGTTTT TCATCCTGCC GCCGCTGGCG GGGATGCTGG TTCAGCCGAT
     301 AGTGGCTACT ACTCAGACCG CACTTGGAAAG CCGCGCTTGG GCGGCCGCGC
     351 CCTGCCGTAT CTGCTTTACG GCACGCTGAT TCGCGTCATC GTGATGATTT
     401 TGATGCCGAA CTCGGGCAGC TTCGGTTTCG GCTATGCGTC GCTGGCGGCC
     451 TTGTCGTTTC GCGCGCTGAT GATTGCGCTG TTGGACGTGT CGTCGAATAT
     501 GGCGATGCAG CCGTTTAAAG TGATGGTCGG CGATATGGTC AACGAGGAGC
     551 AGAAAAGCTA CGCCTACGGG ATTCAAAGTT TCTTAGCGAA TACGGACGCG
     601 GTTGTGGCAG CGATTCTGCC GTTTGTGTTT CCGTATATCG GTTTGGCGAA
     651 CACTGCCGAG AAAGGCGTTG TGCCACAAAC CGTGGTCGTA GCATTCTATG
     701 TGGGTGCGGC GTTACTGATT ATTACCAAGT CGTTCACAAT CTCCAAAGTC
     751 AAAGAATACG ACCCGGAAC CTACGCCCGT TACCACGGCA TCGATGTCGC
     801 CGCGAATCAG GAAAAGCCA ACTGGTTCGA ACTCTTAAAA ACCGCGCCTA
     851 AAGTGTTTTG GACGGTTACT CCGGTACAGT TTTTCTGCTG GTTCGCCTTC
     901 CCGTATATGT GGAATTACTC GGCAGGCGCG ATTGACAGAA ACGTCTGGCA
     951 CACTACCGAT GCGTCTTCCG TAGGCCATCA GGAGCGGGC AACCAGTACG
    1001 GCGTTTGGC GCGGTGTAG

```

This encodes a protein having amino acid sequence <SEQ ID 154>:

```

      1 MIGDRRAGNH FGFSKANTFQ IKKKDLLYVG IYASNSKTRF ARAGKKHDL
     51 VELRLSRRSD GLYPAKLADE PHFSNARRRP AQFGLVFHFA AAGGDAGSAD
     101 SGYSDRTWK PRLGGRRLLPY LLYGTLIIVI VMILMPNSGS FGFGYASLAA
     151 LSFALMIAL LDVSSNMAMQ PFKMMVGDMV NEEQKSYAYG IQSFLANTDA
     201 VVAAILPFVF AYIGLANTAE KGVVPQTVVV AFYVGAALLI ITSFTISKV
     251 KEYDPETYAR YHGIDVAANQ EKANWFELLK TAPKVFVTVT PVQFFCWFAF
     301 RYMWYTYSAGA IAENVWHTTD ASSVGHQEAG NRYGVLAIV*

```

50 ORF16ng and ORF16-1 show 89.3% identity in 261 aa overlap:

```

      30      40      50      60      70      80
      orf16-1.pep  MLSFGLGVQTAFTLQSSQMSRIFQTLGADPHNLGWFILPPLAGMLVQPI-VGHYSDRT
      orf16ng      DVELRLSRRSDGLYPAKLADEPHFSNARRRPAQFGLVF-HPAAAGGDAGSADSGYSDRT
     55      50      60      70      80      90      100
      orf16-1.pep  WKPRLGGRRLLPYLLYGTLIIVIMILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMA
      orf16ng      WKPRLGGRRLLPYLLYGTLIIVIMILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMA
     60      110      120      130      140      150      160

```

10

15

20

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 155>:

25

30

35

Further work revealed the complete nucleotide sequence <SEQ ID 157>:

40

45

50

55

1	MLFRKTTAAV	LAAT <del>L</del> MLNGC	TLMWGMMNP	VSETITRKHV	DKDQIRAFGV
51	VAEDNAQLEK	GSLVMMGGKY	WFVVPEDSA	KLTGILKAGL	DKPFQIVEDT
101	PSYARHQALP	VKLESPGSON	FSTEGLCRLY	DTDKPADIAK	LKQLGFEAVK
151	LDNRTIYTRC	VSAKGKYYAT	PQKLNADYHF	EQSVPADIYY	TVTEEHTDKS

### Homology with a predicted ORF from *N.meningitidis* (strain A)

5 *meningitidis:*

20 The complete length ORF28a nucleotide sequence <SEQ ID 159> is:

This encodes a protein having amino acid sequence <SEQ ID 160>:

**ORF28a and ORF28-1 show 86.1% identity in 238 aa overlap:**

			10	20	30	40	50	60
45	orf28a.pep		MLFRKTTAAVLAATLMLNGCTVMMWGMS	PFSETTARKHVDKDKQIRAFGVVAEDNAQLEK				
	orf28-1		MLFRKTTAAVLAATLMLNGCTLMLWGMN	PNVSETITRKHVDKDKQIRAFGVVAEDNAQLEK				
			10	20	30	40	50	60
50	orf28a.pep		70	80	90	100	110	119
			GSLVMMGGKYWFVVPEDSAKLTGILKAGLDKQFQ	MVPEPNRFA-YQALPVKLESPASQN				
	orf28-1		GSLVMMGGKYWFVVPEDSAKLTGILKAGLDKPFQ	IVEDTPSYARHQALPVKLESPGQN				
			70	80	90	100	110	120
55	orf28a.pep	120	130	140	150	160	170	179
			FSTEGLCRLRYDTRDPADIAKLGKQLEFEAVEL	DNRTIYTRCVSAKGKYYATPQKLNADYHF				
	orf28-1		FSTEGLCRLRYDTPADIAKLGKQLEFEAVEL	DNRTIYTRCVSAKGKYYATPQKLNADYHF				
			130	140	150	160	170	180

	180	190	200	210	220	230
orf28a.pep	EQSVPAIIYYTVTKKHTDKSKLFENIAYTPTTLILDAVGAVLALPVAALIAATNSSDKX					
	:     :     :     :     :     :     :					
5 orf28-1	EQSVPAIIYYTVTEEHTDKSKLFANILYTPPFLILDAAGAVLALPAAALGAVVDAARKX					
	190	200	210	220	230	

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF28 shows 84.2% identity over a 120aa overlap with a predicted ORF (ORF28.ng) from *N.*

#### 10 *gonorrhoeae*:

orf28.pep	MLFRKTTAAVLAHTLMLNGCTLMLWGMNPNVSETITRKHVXKDQIRXFGVVAEDNAQLEK	60
	:     :     :     :     :     :     :	
orf28ng	MLFRKTTAAVLAATLILNGCTMMLRGMNPNVSTITRKHVDKDQIRAFGVVAEDNAQLEK	60
15 orf28.pep	GSLVMMGGKYWFVNPEDSAXXTGILXAGLDKPFQIVXDTPSYXCHQALPVKLGSGXGSON	120
	:     :     :     :     :     :     :	
orf28ng	GSLVMMGGKYWFAVNPEDSAKLTGLKAGLDKPFQIVEDTPSYARHQALPVKFEAPGSON	120

The complete length ORF28ng nucleotide sequence <SEQ ID 161> is

20	1	ATGTTGTTC	GTAAACGAC	CGCCGCCGT	TTGGCGGCA	CCTTGATACT
	51	GAACGGCTGT	ACGATGATGT	TGCGGGGGAT	GAACAACCCG	GTCAGCCAAA
	101	CAATCACCCG	CAAACACGTT	GACAAAGACC	AAATCCGCGC	CTTCGGTGTG
	151	GTTGCCGAAG	ACAATGCCCA	ATTGGAAGAG	GGCAGCCTGG	TGATGATGGG
	201	CGGGAATAC	TGTTTCGCCG	TCAATCCCGA	AGATTCGGCG	AAGCTGACGG
	251	GCCTTTTGAA	GGCCGGGTG	GACAAGCCCT	TCCAAATAGT	TGAGGATACC
25	301	CCGAGCTATG	CCCGCCACCA	AGCCCTGCCG	GTCAAATTCG	AAGCGCCCGG
	351	CAGCCAGAAT	TTCAGTACCG	GAGGTCTTTG	CCTGCGCTAT	GATACCCGCA
	401	GACCTGACGA	CATCGCCAAG	CTGAACACAG	TTGAGTTTAA	AGCGGTCAAA
	451	CTCGACAATC	GGACCATTTA	CACGCGCTGC	GTATCCGCCA	AAGGCAATA
	501	CTACGCCACG	CGCAAAAAC	TGAACGCCGA	TTATCATTTT	GAGCAAAGTG
30	551	TGCCCCCGGA	TATTTATTAT	ACGGTTACTG	AAAAACATAC	CGACAAATCC
	601	AAGCTGTTTG	GAAATATCTT	ATATACGCC	CCCTTGTTGA	TATTGGATGC
	651	GGCGGCCGCG	GTGCTGGTCT	TGCCTATGGC	TCTGATTGCA	GCCGCGAATT
	701	CCTCAGACAA	ATGA			

This encodes a protein having amino acid sequence <SEQ ID 162>:

35	1	MLFRKTTAAV	LAATLILNGC	TMMLRGMNPN	VSQTITRKHV	DKDQIRAFGV
	51	VAEDNAQLEK	GSLVMMGGKY	WFAVNPEDSA	KLTGLLKAGL	DKPFQIVEDT
	101	PSYARHQALP	VKFEAPGSON	FSTGGLCLRY	DTGRPDIAK	LKQLEFKAVK
	151	LDNRTIYTRC	VSAKGKYYAT	PQKLNADYHF	EQSVPAIIYY	TVTEKHTDKS
	201	KLFGNILYTP	PLILDAAAA	VLVLPMALIA	AANSSDK*	

#### 40 ORF28ng and ORF28-1 share 90.0% identity in 231 aa overlap:

	10	20	30	40	50	60
orf28-1.pep	MLFRKTTAAVLAATLMLNGCTLMLWGMNPNVSETITRKHVDKDQIRAFGVVAEDNAQLEK					
	:     :     :     :     :     :					
45 orf28ng	MLFRKTTAAVLAATLILNGCTMMLRGMNPNVSTITRKHVDKDQIRAFGVVAEDNAQLEK					
	10	20	30	40	50	60
	70	80	90	100	110	120
orf28-1.pep	GSLVMMGGKYWFVNPEDSAKLTGILKAGLDKPFQIVEDTPSYARHQALPVKLESPGSON					
	:     :     :     :     :     :					
50 orf28ng	GSLVMMGGKYWFAVNPEDSAKLTGLKAGLDKPFQIVEDTPSYARHQALPVKFEAPGSON					
	70	80	90	100	110	120
	130	140	150	160	170	180
orf28-1.pep	FSTEGCLRYDTPKPADIAKLKQLGFEAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF					
	:         :     :     :     :     :     :					
55 orf28ng	FSTGGLCLRYDTPGRPDIAKLKQLEFKAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF					
	130	140	150	160	170	180
	190	200	210	220	230	239
orf28-1.pep	EQSVPAIIYYTVTEEHTDKSKLFANILYTPPFLILDAAGAVLALPAAALGAVVDAARKX					
	:     :     :     :     :     :					
60 orf28ng	EQSVPAIIYYTVTEKHTDKSKLFGNILYTPPLILDAAAVVLVLPMALIAAANSSDKX					

190 200 210 220 230

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF28-1 (24kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 6A shows the results of affinity purification of the GST-fusion protein, and Figure 6B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result. These experiments confirm that ORF28-1 is a surface-exposed protein, and that it may be a useful immunogen.

#### Example 20

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 163>:

```

1  ..GTCAGTCCTG TACTGCCTAT TACACACGAA CGGACAGGGT TTGAAGGTGT
15 51 TATCGGTTAT GAAACCCATT TTTCAGGGCA CGGACATGAA GTACACAGTC
101 CGTTCGATCA TCATGATTCA AAAAGCACTT CTGATTTTCTAG CGGCGGTGTA
151 GACGGCGGTT TTACTGTTTA CCAACTTCAT CGAACATGGT CGGAAATCCA
201 TCCGGAGGAT GAATATGACG GGCCGCAAGC AGCG.ATTAT CCGCCCCCCG
251 GAGGAGCAAG GGATATATAC AGCTATTATG TCAAAGGAAC TTCAACAAAA
20 301 ACAAAGACTA GTATTGTCCC TCAAGCCCCA TTTTCAGACC GTTGGCTAGA
351 AGAAAATGCC GGTCCCGCCT CTGGT..

```

This corresponds to the amino acid sequence <SEQ ID 164; ORF29>:

```

1  ..VSPVLPITHE RTGFEGVIGY ETHFSGHGHE VHSPFDHDS KSTSDFSGGV
25 51 DGGFTVYQLH RTWSEIHPED EYDGPQAAXY PPPGGARDIY SYVVKGTSTK
101 TKTSIVPQAP FSDRWLEENA GAASG..

```

Further work revealed the complete nucleotide sequence <SEQ ID 165>:

```

1  ATGAATTTGC CTATTCAAAA ATTCATGATG CTGTTTGAGC CAGCAATATC
51 GTTGTGCAA ATCCCCATTA GTCATGCGAA CGGTTTGGAT GCCCGTTTGC
101 GCGATGATAT GCAGGCAAAA CACTACGAAC CGGGTGGTAA ATACCATCTG
30 151 TTTGGTAATG CTCGCGGCAG TGTAAAAAAG CGGGTTTACG CCGTCCAGAC
201 ATTTGATGCA ACTGCGGTCA GTCCTGTACT GCCTATTACA CACGAACGGA
251 CAGGGTTTGA AGGTGTTATC GGTATGAAA CCCATTTTTC AGGGCAGCGA
301 CATGAAGTAC ACAGTCCGTT CGATCATCAT GATTCAAAAA GCACTTCTGA
351 TTTCAGCGGC GGTGTAGACG GCGGTTTAC TGTTTACCAA CTTTCATCGA
35 401 CAGGTTCGGA AATCCATCCG GAGGATGGAT ATGACGGGCC GCAAGGCAGC
451 GATTATCCGC CCCCCGGAGG AGCAAGGGAT ATATACAGCT ATTATGTCAA
501 AGGAACTTCA ACAAAAACAA AGACTAATAT TGTCCCTCAA GCCCCATTTT
551 CAGACCGTTG GCTAAAAGAA AATGCCGGTG CCGCCTCTCG TTTTTCAGC
601 CGTGCAGATG AAGCAGGAAA ACTGATATGG GAAAGCGACC CCAATAAAAA
40 651 TTGGTGGGCT AACCGTATGG ATGATGTTTC CGGCATCGTC CAAGGTGCGG
701 TTAATCCTTT TTAATGGGT TTTCAAGGAG TAGGGATTGG GGCAATTACA
751 GACAGTGACG TAAGCCCGGT CACAGATACA GCCGCGCAGC AGACTCTACA
801 AGGTATTAAT GATTAGGAA AATTAAGTCC GGAAGCACAA CTTGCTGCCG
851 CGAGCCTATT ACAGGACAGT GCTTTTGGCG TAAAAGACGG TATCAACTCT
45 901 GCCAAACAAT GGGCTGATGC CCATCCAAAT ATAACAGCTA CTGCCCAAAC
951 TGCCCTTTCC GCAGCAGAGG CCGCAGGTAC GGTTTGGAGA GGTAATAAAG
1001 TAGAAGTTAA CCGACTAAA TGGGATTGGG TTAAAAATAC CGGTTATAAA
1051 AAACCTGCTG CCCGCCATAT GCAGACTTTA GATGGGGAGA TGGCAGGTGG
1101 GAATAAACCT ATTAAATCTT TACCAAACAG TGCCGCTGAA AAAAGAAAAC
50 1151 AAAATTTTGA GAAGTTTAAT AGTAACTGGA GTTCAGCAAG TTTTGATTCA

```

-144-

1201 GTGCACAAAA CACTAACTCC CAATGCACCT GGTATTTTAA GTCCTGATAA  
 1251 AGTTAAAACT CGATACACTA GTTTAGATGG AAAAATTACA ATTATAAAG  
 1301 ATAACGAAAA CAACTATTTT AGAATCCATG ATAATTCACG AAAACAGTAT  
 1351 CTTGATTCAA ATGGTAATGC TGTGAAAACC GGTAAATTAC AAGGTAAGCA  
 5 1401 AGCAAAAGAT TATTTACAAC AACAACTCA TATCAGGAAC TTAGACAAAT  
 1451 GA

This corresponds to the amino acid sequence <SEQ ID 166; ORF29-1>:

1 MNLPIQKFM LFAAAISLLQ IPISHANGLD ARLRDMQAK HYPEGGKYHL  
 51 FGNARGSVKK RVYAVQTFDA TAVSPVLPIT HERTGFEGVI GYETHFSGHG  
 10 101 HEVHSPFDHH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP EDGYDGPQGS  
 151 DYPPPGGARD IYSYVKGTS TKTKTNIVPQ APFSDRWLKE NAGAASGFFS  
 201 RADEAGKLIW ESDPNKNWWA NRMDDVRGIV QGAVNPFLMG FQGVGIGAIT  
 251 DSAVSPVTD AAQOTLQGIN DLGKLSPEAQ LAAASLLQDS AFAVKDGINS  
 301 AKQWADAHNP ITATAQTALS AAEEAGTVWR GKKVELNPTK WDWVKNTGYK  
 15 351 KPAARHMOTL DGEMAGGNKP IKSLPNSAAE KRKQNFKEFN SNWSSASFDS  
 401 VHKTLPNAP GILSPDKVKT RYTSLDGKIT IIKDNENNYF RIHDNSRKQY  
 451 LDSNGNAVKT GNLQGGKQAD YLQQQTHIRN LDK\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

20 ORF29 shows 88.0% identity over a 125aa overlap with an ORF (ORF29a) from strain A of *N. meningitidis*:

					10	20	30
	orf29.pep				VSPVLPIT	HERTGFEGVI	GYETHFSGHGHE
					: :	: :	: :
25	orf29a	EPGGKYHL	FGNARGSVKN	RVYAVQTFDA	TAVGPILP	ITHERTGFEG	IIGYETHFSGHGHE
		50	60	70	80	90	100
		40	50	60	70	80	90
30	orf29.pep	VHSPFDHHD	SKSTSDFSG	GVDGGFTVYQL	HRTWSEIHP	EDGYDGPQA	AXYPPPGGARDIY
		: :	: :	: :	: :	: :	: :
	orf29a	VHSPFDNHD	SKSTSDFSG	GVDGGFTVYQL	HRTGSEIHP	EDGYDGPQGS	DYPPPGGARDIY
		110	120	130	140	150	160
		100	110	120			
35	orf29.pep	SYVVKGTST	TKTKTSIVP	QAPFSDRW	LEENAGAASG		
		: :	: :	: :	: :	: :	: :
	orf29a	XXYVVKGT	STTKTSNIV	PRAPFSDR	WLKENAGA	ASGFFSRA	DEAGKLIWESDPNKNWWANR
		170	180	190	200	210	220
40	orf29a	MDDIRGIV	QGAVNPFL	MGFQGVGIG	AITDSAVS	PVTDAAQ	OTLQGNHLGXLSPEAQLA
		230	240	250	260	270	280

The complete length ORF29a nucleotide sequence <SEQ ID 167> is:

1 ATGAATTNGC CTATTCAAAA ATTCATGATG CTGTTTGCGAG CAGCAATATC  
 45 51 GTNGCTGCAA ATCCCNATTA GTCATGCGAA CGGTTTGGAT GCCCGTTTGC  
 101 GCGATGATAT GCAGGCAAAA CACTACGAAC CGGGTGGTAA ATACCATCTG  
 151 TTTGGTAATG CTCGCGGCAG TGTTAAAAAT CGGGTTTACG CCGTCCAAC  
 201 ATTTGATGCA ACTGCGGTCG GCCCCATACT GCCTATTACA CACGAACGGA  
 251 CAGGATTGTA AGGCATTATC GGTATGAAA CCCATTTTTC AGGACATGGA  
 301 CATGAAGTAC ACAGTCCGTT CGATAATCAT GATTCAAAAA GCACTTCTGA  
 50 351 TTTACGCGGC GCGGTAGACG GTGGTTTAC CGTTACCAA CTTATCGGA  
 401 CAGGGTCGGA AATCCATCCG GAGGATGGAT ATGACGGGCC GCAAGGCAGC  
 451 GATTATCCGC CCCCAGGAG AGCAAGGAT ATATACANNT ANTATGTCAA  
 501 AGGAACCTCA ACAAAAACAA AGAGTAATAT TGTCCCCGA GCCCATTTT  
 551 CAGACCGCTG GCTAAAAGAA AATGCCGGTG CCGCTCTGG TTTTTCAGC  
 55 601 CGTGCTGATG AAGCAGGAAA ACTGATATGG GAAAGCGACC CCAATAAAAA  
 651 TTGGTGGGCT AACCGTATGG ATGATATTCG CGGCATCGTC CAAGGTGCGG  
 701 TTAATCCTTT TTAATGGGT TTTCAAGGAG TAGGGATTGG GGCAATTACA  
 751 GACAGTGCGAG TAAGCCCGGT CACAGATACA GCCGCGCAGC AGACTCTACA  
 801 AGGTATNAAT CATTTAGGAA ANTTAAGTCC CGAAGCACAA CTTGCGGCTG  
 60 851 CAACCGCATT ACAAGACAGT GCTTTTGCGG TAAAAGACGG TATCAATTCC  
 901 GCCAGACAAT GGGCTGATGC CCATCCGAAT ATAACGCAA CAGCCCAAAC



-145-

951 TGCCCTTGCC GTAGCAGANG CCGCAACTAC GGTTTGGGGC GGTAAGAAAG  
 1001 TAGAAGTTAA CCCGACCAAA TGGGATTGGG TTAAGAAATAC NGGCTATAAN  
 1051 ACACCTGCTG TTCGCACCAT GCATACTTTG GATGGGGAAA TGGCCGGTGG  
 1101 GAATAGACCG CCTAAATCTA TAACGTCCAA CAGCAAAGCA GATGCTTCCA  
 1151 CACAACCGTC TTTACAAGCG CAACATAATTG GAGAACAAT TANNNNNGGG  
 1201 CATGCTTATA ACAAGCATGT CATAAGACAA CAAGAATTTA CGGATTAAAA  
 1251 TATCAATTCA CCAGCAGATT TTGCTCGGCA TATTGAAAT ATTGTAGCC  
 1301 ATCCANCAAA TATGAAAGAG TTACCTCGCG GTAGAACTGC GTATTGGGAT  
 1351 NATAAACAG GGACNATAGT TATCCGAGAT AAAAATTCTG ACGATGGAGG  
 1401 TACAGCATT AGACCAACAT CAGGTAAAA ATATTATGAT GATTATAG

This encodes a protein having amino acid sequence <SEQ ID 168>:

1 MNXPIQKFM LFAAISXLQ IPISHANGLD ARLRDDMOAK HYPEGGKYHL  
 51 FGNARGSVKN RVYAVQTFDA TAVGPILPIT HERTGFEGII GYETHFSGHG  
 101 HEVHSPFDNH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP EDGYDGPQGS  
 151 DYPPPGGARD IYXXYVKGTS TKTKSNIVPR APFSDRWLKE NAGAASGFFS  
 201 RADEAGKLIW ESDPNKNWWA NRMDDIRGIV QGAVNPFLMG FQGVGIGAIT  
 251 DSAVSPVTD AAQOTLQGXN HLGXLSPEAQ LAAATALQDS AFAVKDGIN  
 301 ARQWADAHN ITATAQTALA VAXAATTWVG GKKVELNPTK WDWKNTGYX  
 351 TPAVRTMHTL DGEMAGGNRP PKSITSNSKA DASTQPSLQA QLIGEIXXG  
 401 HAYNKHVIRQ QEFTDLNINS PADFARHIEN IVSHPNXMK LPRGRTAYWD  
 451 XKTGTIVIRD KNSDDGGTAF RPTSGKKYYD DL\*

ORF29a and ORF29-1 show 90.1% identity in 385 aa overlap:

		10	20	30	40	50	60
25	orf29a.pep	MNXPIQKFM LFAAISXLQ IPISHANGLD ARLRDDMOAK HYPEGGKYHL FGNARGSVKN					
	orf29-1	MNLPIQKFM LFAAISLLQ IPISHANGLD ARLRDDMOAK HYPEGGKYHL FGNARGSVKK					
		10	20	30	40	50	60
30	orf29a.pep	RVYAVQTFDA TAVGPILPIT HERTGFEGII GYETHFSGHG HEVHSPFDNH DSKSTSDFSG					
	orf29-1	RVYAVQTFDA TAVSPVLPIT HERTGFEGVIGYETHFSGHG HEVHSPFDNH DSKSTSDFSG					
		70	80	90	100	110	120
35	orf29a.pep	GVDGGFTVYQLHRTGSEIHPEDGYDGPQGS DYPPPGGARDIYXXYVKGTS TKTKSNIVPR					
	orf29-1	GVDGGFTVYQLHRTGSEIHPEDGYDGPQGS DYPPPGGARDIYXXYVKGTS TKTKSNIVPQ					
		130	140	150	160	170	180
40	orf29a.pep	APFSDRWLKENAGAASGFFSRADEAGKLIWESDPNKNWWANRMDDIRGIVQGAVNPFLMG					
	orf29-1	APFSDRWLKENAGAASGFFSRADEAGKLIWESDPNKNWWANRMDDVRGIVQGAVNPFLMG					
		190	200	210	220	230	240
45	orf29a.pep	FQGVGIGAITDSAVSPVTDAAQOTLQGXNHLGXLSPEAQ LAAATALQDS AFAVKDGIN					
	orf29-1	FQGVGIGAITDSAVSPVTDAAQOTLQGXNHLGXLSPEAQ LAAATALQDS AFAVKDGIN					
		250	260	270	280	290	300
50	orf29a.pep	ARQWADAHNITATAQTALAVAXAATTWVG GKKVELNPTKWDWKNTGYXTPAVRTMHTL					
	orf29-1	AKQWADAHNITATAQTALAAEAGTVWRGKKVELNPTKWDWKNTGYKPPAARHMTL					
		310	320	330	340	350	360
55	orf29a.pep	DGEMAGGNRPKPKSITSNSKADASTQPSLQAQLIGEIXXGHAYNKHVIRQ QEFTDLNINS					
	orf29-1	DGEMAGGNRPKPKSLP-NSAAEKRKQNFENSNWSSASFDVHKTLPNAPGILSPDKVK					
		370	380	390	400	410	420
60	orf29a.pep						
	orf29-1						

Homology with a predicted ORF from *N.gonorrhoeae*

ORF29 shows 88.8% identity over a 125aa overlap with a predicted ORF (ORF29.ng) from *N. gonorrhoeae*:

```

5      orf29.pep                      VSPVLPITHERTGFEGVIGYETHFSGHGHE 30
      orf29ng      EPGGKYHLFGNARGSVKNRVCVQTFDATAVGPILPITHERTGFEGVIGYETHFSGHGHE 102
10     orf29.pep      VHSFPDHHDSKSTSDFSGGVDGGFTVYQLHRTWSEIHPEDYDGPQAAAXYPPPGGARDIY 90
      orf29ng      VHSFPDNHDSKSTSDFSGGVDGGFTVYQLHRTGSEIHPEDGYDGPQGGGYPPPGGARDIY 162
      orf29.pep      SYVVKGTSTKTKTSIVPQAPFSDRWLEENAGAASG 125
      orf29ng      SYHIKGTSTKTKINTVPQAPFSDRWLKENAGAASGFLSRADEAGKLIWENDPDKNWRANR 222

```

15 The complete length ORF29ng nucleotide sequence <SEQ ID 169> is predicted to encode a protein having amino acid sequence <SEQ ID 170>:

```

1      MNLPIQKFMM LFAAAISLLQ IPISHANGLD ARLRDDMQAK HYEPPGGKYHL
51     FGNARGSVKN RVCVQTFDA TAVGPILPIT HERTGFEGVI GYETHFSGHG
101    HEVHSPFDNH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP EDGYDGPQGG
151    GYPPPGGARD IYSYHIKGT TTKKINTVPQ APFSDRWLKE NAGAASGFLS
201    RADEAGKLIW ENDPDKNWR NRMDDIRGIV QGAVNPFLTG FQGLGVGAIT
251    DSAVSPVTYA AARKTLQGIH NLGNLSPEAQ LAAATALQDS AFAVKDSINS
301    ARQWADAHFN ITATAQTALA VTEAATTWVG GKVELNPAK WDWVKNTGYK
351    KPAARHMQTV DGEMAGGNKP LESKNTVTN NFFENTGYTE KVLQASNGD
251    YHGFPPQSVDA FSENGTVIQI VGGDNIVRHK LYIPGSYKKG DGNFEYIREA
451    DGKINHRLFV PNQQLPEK*

```

In a second experiment, the following DNA sequence <SEQ ID 171> was identified:

```

1      atgAATTTGC CTATTCAAAA ATTCATGATG ctgttggcAg cggcaatatc
51     gatgctGCat ATCCCCATTA GTCATGCGAA CGGTTTGAT GCCCGTTTGC
30     GCGATGATAT GCAGGCAAAA CACTACGAAC CGGGTGGCAA ATACCATCTG
151    TTTGGTAATG CTCGCGGCAG TGTTAAAAAT CGGTTTGCG CCGTCCAAAC
201    ATTTGATGCA ACTGCGGTCG GCCCCATACT GCCTATTACA CACGAACGGA
251    CAGGATTTGA AGGTGTTATC GGCTATGAAA CCCATTTTTC AGGACACGGA
301    CACGAAGTAC ACGTCCGTT CGATAATCAT GATTCAAAAA GCACCTCTGA
35     TTTAGCGGC GCGGTAGACG GCGGTTTAC CGTTTACCA CTTATCCGGA
401    CAGGGTCGGA AATACATCCC GCAGACGGAT ATGACGGGCC TCAAGGCGGC
451    GGTTATCCGG AACCACAAGG GGCAAGGGAT ATATACAGCT ACCATATCAA
501    AGGAACCTCA ACCAAAACAA AGATAAACAC TGTTCCGCAA GCCCCTTTT
551    CAGACCGCTG GCTAAAAGAA AATGCCGGTG CCGCTTCCGG TTTTCTCAGC
40     CGTCCGATG AAGCAGGAAA ACTGATATGG GAAACGACC CCGATAAAAA
601    TTGGCGGGCT AACCGTATGG ATGATATTCG CGGCATCGTC CAAGGTGCGG
701    TTAATCCTTT TTTAACGGGT TTTCAAGGGG TAGGGATTGG GGCAATTACA
751    GACAGTGGCG TAAGCCCGGT CACAGATACA GCCGCTCAGC AGACTCTACA
801    AGGTATTAAT GATTTAGGAA ATTTAAGTCC GGAAGCACAA CTTGCCCGCG
45     CGAGCCTATT ACAGGACAGT GCCTTTGCGG TAAAAGACGG CATCAATTCC
901    GCCAGACAAT GGGCTGATGC CCATCCGAAT ATAACAGCAA CAGCCCAAAC
951    TGCCCTTGCC GTAGCAGAGG CCGCAGGTAC GGTTTGGCGC GGTAAAAAAG
1001   TAGAACTTAA CCCGACCAA TGGGATTGGG TTTAAAATAC CGGCTATAAA
1051   AAACCTGCTG CCCGCCATAT GCAGACTGTA GATGGGGAGA TGGCAGGGGG
50     GAATAGACCG CCTAAATCTA TAACGTCGGA AGGAAAAGCT AATGCTGCAA
1151   CCTATCCTAA GTTGGTTAAT CAGCTAAATG AGCAAAACTT AAATAACATT
1201   GCGGCTCAAG ATCCAAGATT GAGTCTAGCT ATTCATGAGG GTAAAAAAA
1251   TTTTCCAATA GGAAGTCAA CTTATGAAGA GGCAGATAGA CTAGGTAAA
1301   TTTGGGTTGG TGAGGGTGCA AGACAACTA GTGGAGGCGG ATGGTTAAGT
55     AGAGATGGCA CTCGACAATA TCGGCCACCA ACAGAAAAAA AATCACAATT
1401   TGCAACTACA GGTATTCAAG CAAATTTTGA AACTTATACT ATTGATTCAA
1451   ATGAAAAAAG AAATAAAATT AAAAATGGAC ATTTAAATAT TAGGTAA

```

This encodes a protein having amino acid sequence <SEQ ID 172; ORF29ng-1>:

```

60     1      MNLPIQKFMM LLAAISMILH IPISHANGLD ARLRDDMQAK HYEPPGGKYHL
      51     FGNARGSVKN RVCVQTFDA TAVGPILPIT HERTGFEGVI GYETHFSGHG

```

101 HEVHSPFDNH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP ADGYDGPQGG  
 151 GYPEPQGARD IYSYHIKGTSTKTKINTVPQ APFSDRWLKE NAGAASGFSL  
 201 RADEAGKLIW ENDPKKNWRA NRMDDIRGIV QGAVNPFLTG FQGVGIGAIT  
 251 DSAVSPVTD AAQQTLOGIN DLGNLSPEAQ LAAASLLQDS AFAVKDGIN  
 301 ARQWADAHNP ITATAQTALA VAEAAGTVWR GKKVELNPTK WDWVKNTGYK  
 351 KPAARHMQTV DGEMAGGNRP PKSITSEGKA NAATYPKLVN QLNEQNLNNI  
 401 AAQDPRLSLA IHEGKKNFPI GTATYEEADR LGKIWVGEGA RQTSGGGWLS  
 451 RDGTRQYRPP TEKKSQFATT GIQANFETYT IDSNEKRNKI KNGHLNIR\*

ORF29ng-1 and ORF29-1 show 86.0% identity in 401 aa overlap:

10	orf29ng-1.pep	10	20	30	40	50	60
	orf29-1	10	20	30	40	50	60
15	orf29ng-1.pep	70	80	90	100	110	120
	orf29-1	70	80	90	100	110	120
20	orf29ng-1.pep	130	140	150	160	170	180
	orf29-1	130	140	150	160	170	180
25	orf29ng-1.pep	190	200	210	220	230	240
	orf29-1	190	200	210	220	230	240
30	orf29ng-1.pep	250	260	270	280	290	300
	orf29-1	250	260	270	280	290	300
35	orf29ng-1.pep	310	320	330	340	350	360
	orf29-1	310	320	330	340	350	360
40	orf29ng-1.pep	370	380	390	400	410	419
	orf29-1	370	380	390	400	410	420
45	orf29ng-1.pep	420	430	440	450	460	479
	orf29-1	420	430	440	450	460	480

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 21

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 173>:

-148-

1 ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC  
 51 CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAATGTTCC  
 101 ACACGCGGGC AGATGCACCG ATGCAG...

This corresponds to the amino acid sequence <SEQ ID 174; ORF30>:

5 1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QMFHTRADAP MQ..

Further work revealed the complete nucleotide sequence <SEQ ID 175>:

10 1 ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC  
 51 CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAGTGTTCC  
 101 ACACGCGGGC AGATGCACCG ATGCAGTTGG CGGAGCTTTC TCAAAAGGAG  
 151 ATGAAGGAGA CAGAGGGGGC GTTCTCTCCA TTGGCTATCT TGGGTGGTGC  
 201 TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA  
 251 GACCAGCTTC TGTTAGAGAT GTTGCTATTG CTGGCGGATT AGGCGCAATT  
 301 CCTGGTGGTG TAGGCGCCGC AGGAAAGGTT GTTTCCTTTG CTAAATATGG  
 351 ACGTGAGATT AAAATCGGCA ATAATATGCG GATAGCCCTT TTCGGTAATA  
 15 401 GAACAGGTCA TCCTATTGGA AAATTTCCTT ATTATCATCG TCGAGTTACG  
 451 GATAATACGG GCAAGACTTT GCCTGGACAG GGAATTGGTC GTCATCGCCC  
 501 TTGGGAATCA AAATCTACGG ACAGATCATG GAAAACCCG TTCTAA

This corresponds to the amino acid sequence <SEQ ID 176; ORF30-1>:

20 1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE  
 51 MKETEGAFLEP LAILGGAAIG MWTQHGFSA TTGRPASVRD VAIAGGLGAI  
 101 PGGVGAAGKV VSAKYGREI KIGNNMRIAP FGNRTGHPIG KFPYHRRVT  
 151 DNTGKTLPGQ GIGRHRPWES KSTDRSWKNR F\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF30 shows 97.6% identity over a 42aa overlap with an ORF (ORF30a) from strain A of *N. meningitidis*:

	10	20	30	40
orf30.pep	MKKQITAAVMMLSMIAPAMANGLDNQAFEDQMFHTRADAPMQ			
30 orf30a	MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKXTXGAFLP			
	10	20	30	40
	50	60		
orf30a	LXILGGAAIGMWTQHGFSAATTGRPASVRDVAIAGGLGAIPGXVGAAGKVVSFAKYGREI			
	70	80	90	100
	110	120		

35 The complete length ORF30a nucleotide sequence <SEQ ID 177> is:

1 ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC  
 51 CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAGTGTTCC  
 101 ACACGCGGGC AGATGCACCG ATGCAGTTGG CGGAGCTTTC TCAAAAGGAG  
 40 151 ATGAAGGANA CAGNGGGGGC GTTCTCTCCA TTGGNTATCT TGGGTGGTGC  
 201 TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA  
 251 GACCAGCTTC TGTTAGAGAT GTTGCTATTG CTGGCGGATT AGGCGCAATT  
 301 CCTGGTGNTG TAGGCGCCGC AGGAAAGGTT GTTTCCTTTG CTAAATATGG  
 351 ACGTGAGATT AAAATCGGCA ATAATATGCG GATAGCCCTT TTCGGTAATA  
 401 GAACAGGTCA TCCTATTGGA AAATTTCCTT ATTATCATCG TCGAGTTACG  
 45 451 GATAATACGG GCAAGACTTT GCCTGGACAG GGAATTGGTC GTCATCGCCC  
 501 TTGGGAATCA AAATCTACGG ACAGATCATG GAAAACCCG TTCTAA

This encodes a protein having amino acid sequence <SEQ ID 178>:

50 1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE  
 51 MKXTXGAFLP LXILGGAAIG MWTQHGFSA TTGRPASVRD VAIAGGLGAI  
 101 PGXVGAAGKV VSAKYGREI KIGNNMRIAP FGNRTGHPIG KFPYHRRVT  
 151 DNTGKTLPGQ GIGRHRPWES KSTDRSWKNR F\*

ORF30a and ORF30-1 show 97.8% identity in 181 aa overlap:

orf30a.pep MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKXTXGAFLP 60

	orf30-1	 MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP	60
5	orf30a.pep	LXILGGAAGMWTQHGFYSYATTGRPASVRDVAIAGGLGAI PGXVGAAGKVVSFAKYGREI	120
	orf30-1	LAILGGAAGMWTQHGFYSYATTGRPASVRDVAIAGGLGAI PGGVGAAGKVVSFAKYGREI	120
	orf30a.pep	KIGNNMRIAPFGNRTGHPIGKFPYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR	180
10	orf30-1	KIGNNMRIAPFGNRTGHPIGKFPYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR	180
	orf30a.pep	FX	
15	orf30-1	FX	

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF30 shows 97.6% identity over a 42aa overlap with a predicted ORF (ORF30.ng) from *N.gonorrhoeae*:

20	orf30.pep	MKKQITAAVMMLSMIAPAMANGLDNQAFEDQMFHTRADAPMQ	42
	orf30ng	 MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP	60

The complete length ORF30ng nucleotide sequence <SEQ ID 179> is

	1	ATGAAAAAAC	AAATCACCGC	AGCCGTAATG	ATGCTGTCTA	TGATCGCCCC
25	51	CGCAATGGCA	AACGGATTGG	ACAATCAGGC	ATTGAAGAC	CAAGTGTTC
	101	ACACGCGGGC	AGATGCGCCG	ATGCAGTTGG	CGGAGCTTTC	TCAGAAAGGAG
	151	ATGAAGGAGA	CTGAAGGGGC	TTTTCTTCCA	TTGGCTATCT	TGGGTGGTGC
	201	TGCCATTGGT	ATGTGGACAC	AGCATGGTTT	TAGTTATGCA	ACGACAGGCA
	251	GACCAGCTTC	TGTTAGAGAT	GTTGCTGGCG	GATTAGGCGC	AATTCCTGGT
30	301	GATGTAGGTG	CTGCAGGAAA	GGTTGTTTCC	TTTGCTAAAT	ATGGACGTGA
	351	GATTAAATC	GGCAATAATA	TGCGGATAGC	CCCTTTCGGT	AATAGAACAG
	401	GTCATCCTAT	TGGAAAATT	CCCCATTATC	ATCGTCGAGT	TACGGATAAT
	451	ACGGGCAAGA	CTTTCCTGG	ACAGGGAATT	GGTCGTCATC	GCCCTTGGGA
	501	ATCAAATCT	ACGGACAGAT	CATGAAAAA	CCGCTTCTAA	

This encodes a protein having amino acid sequence <SEQ ID 180>:

35	1	MKKQITAAVM	MLSMIAPAMA	NGLDNQAFED	QVFHTRADAP	MQLAELSQKE
	51	MKETEGAFLP	LAILGGAIG	MWTQHGFYSY	TTGRPASVRD	VAGGLGAIPG
	101	DVGAAGKVVS	FAKYGREIKI	GNNMRIAPFG	NRTGHPIGKF	PHYHRRVTDN
	151	TGKTLPGQGI	GRHRPWESKS	TDRSWKNRF*		

ORF30ng and ORF30-1 show 98.3% identity in 181 aa overlap:

40	orf30ng.pep	10 20 30 40 50 60 MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP
	orf30-1	 MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP
45		10 20 30 40 50 60
	orf30ng.pep	70 80 90 100 110 LAILGGAAGMWTQHGFYSYATTGRPASVRDVA--GGLGAI PGDVGAAGKVVSFAKYGREI
	orf30-1	 LAILGGAAGMWTQHGFYSYATTGRPASVRDVAIAGGLGAI PGGVGAAGKVVSFAKYGREI
50		70 80 90 100 110 120
	orf30ng.pep	120 130 140 150 160 170 KIGNNMRIAPFGNRTGHPIGKFPYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR
55	orf30-1	 KIGNNMRIAPFGNRTGHPIGKFPYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR
		130 140 150 160 170 180
60	orf30ng.pep	180 FX
	orf30-1	FX

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## Example 22

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 181>:

```

1 ATGAATAAAA CTCTCTATCG TGTAATTTTC AACCGCAAAC GTGGGGCTGT
51 GrTAGCCGTT GCTGAAACTA CCAAGCGCGA AGGTAAAAGC TGTGCCGATA
101 GTGATTCAGG CAGCGCTCAT GTGAAATCTG TTCCTTTTGG TACTACTCAT
151 GCACCTGTTT GTg.CGTTaC AAATATCTTT TCTTTTCTT TATTGGGCTT
10 201 TTCTTTATGT TTGGCTGTAG GtacGGyCAA TATTGCTTTT GCTGATGGCA
251 TT..

```

This corresponds to the amino acid sequence <SEQ ID 182; ORF31>:

```

1 MNKTLRVIF NRKRGAVXAV AETTKREGKS CADSDSGSAH VKSVFPGTTH
51 APVCXVTNIF SFSLLGFSLC LAVGTXNIAF ADGI..

```

- 15 Further work revealed a further partial nucleotide sequence <SEQ ID 183>:

```

1 ATGAATAAAA CTCTCTATCG TGTAATTTTC AACCGCAAAC GTGGGGCTGT
51 GGTAGCCGTT GCTGAAACTA CCAAGCGCGA AGGTAAAAGC TGTGCCGATA
101 GTGATTCAGG CAGCGCTCAT GTGAAATCTG TTCCTTTTGG TACTACTCAT
151 GCACCTGTTT GTCGTTCAAA TATCTTTTCT TTTTCTTAT TGGGCTTTC
20 201 TTTATGTTTG GCTGTAGGTA CGGCCAATAT TGCTTTTGCT GATGGCATT..

```

This corresponds to the amino acid sequence <SEQ ID 184; ORF31-1>:

```

1 MNKTLRVIF NRKRGAVVAV AETTKREGKS CADSDSGSAH VKSVFPGTTH
51 APVCRSNIFS FSLLGFSLLC AVGTANIAFA DGI..

```

Computer analysis of this amino acid sequence gave the following results:

- 25 Homology with a predicted ORF from *N.gonorrhoeae*

ORF31 shows 76.2% identity over a 84aa overlap with a predicted ORF (ORF31.ng) from *N.gonorrhoeae*:

```

30 orf31.pep MNKTLRVIFNRKRGAVXAVAETTKREGKSCADSDSGSAHVKSVPFGTTHAPVCXVTNIF 60
orf31ng MNKTLRVIFNRKRGAVVAVAEETTKREGKSCADSGSGSVYVKSVSFIPTH-----SKAF 54
orf31.pep SFSLLGFSLCLAVGTXNIAFADGI 84
orf31ng CFSALGFSLCLALGTVNIAFADGIITDKAAPKTQQATILQTGNGIPQVNIQTPTSAGVSV 114

```

- 35 The complete length ORF31ng nucleotide sequence <SEQ ID 185> is:

```

1 ATGAACAAAA CCCTCTATCG TGTGATTTTC AACCGCAAAC GCGGTGCTGT
51 GGTAGCTGTT GCCGAAACCA CCAAGCGCGA AGGTAAAAGC TGTGCCGATA
101 GTGGTTCGGG CAGCGTTTAT GTGAAATCCG TTTCTTTCAT TCCTACTCAT
151 TCCAAAGCCT TTTGTTTTTC TGCATTAGGC TTTTCTTAT GTTGGGCTTT
40 201 GGGTACGGTC AATATTGCTT TTGCTGACGG CATTATTACT GATAAAGCTG
251 CTCCTAAAAC CCAACAAGCC ACGATTCTGC AAACAGGTaa cGGCATACCG
301 CAAGTCAATA TTCAAACCCC TACTTCGGCA GGGGTTTCTG TTAATCAATA
351 TGCCAGTTT GATGTGGGTA ATCGCGGGGC GATTTTAAAC AACAGTCGCA
401 GCAACACCCA AACACAGCTA GGCGGTTGGA TTCAAGGCAA TCCTTGTTG
45 451 ACAAGGGGCG AAGCACGTGT GGTGTGTAAC CAAATCAACA GCAGCCATCC
501 TTCACAACTG AATGGCTATA TTGAAGTGGG TGGACGACGT GCAGAAGTCG
551 TTATTGCCAA TCCGGCAGGG ATTGCAGTCA ATGGTGGTGG TTTTATCAAT
601 GCTTCCCGTG CCACTTTGAC GACAGGCCAA CCGCAATATC AAGCAGGAGA
651 CTTTAGCGGC TTAAAGATAA GGCAAGGCAA TGCTGTAATC GCCGGACACG

```

1	MNKTLYRVIF	NRKRGAVVAV	AETTKREGKS	CADSGSGSVY	VKSVSFTPTH
51	SKAFCFSAIG	FSLCLALGTV	NIAFADGIIT	DKAAPTQTQA	TLTQTNNGIP
101	QVNIQTPTSA	GSVSNQYQAF	DVGNRGAILN	NSRNPOTQOL	GGWIQGNFWL
151	TRGEARVVVN	QINSHPSQL	NGYIEVGRR	AEVVIANPAG	IAVNGGGFIN
201	ASRATLTGTQ	PQYQAGDFSG	FKIROGNAVI	AGHGLDARDT	DFTRILVCQQ
251	NHLDQYGRTS	RHS*			

orf31ng	96	NGGIPQVNIQTPTSAGVSVNQYAQFDVGNRGAILNNSRSN-TQTQLGGWIOGNPWLTRGE	154
		GNG+P VNI TP ++G+S N+Y F+V NRG ILNN + T +QLGG IQ NP L	
HecA	45	GNGVPPVNIATPDASGLSHNRYHDFNVDRGLILNNGTARLTPTSQGLGLIQNNPNPLNGRA	104
Orf31ng	155	ARVVVNQINSSHPSQLNGYIEVGGRRAEVVIANPAGIAVNGGGFINASRATLTGQPQYQ	214
		A ++N++ S + S+L GY+EV G+ A VV+ANP GI +G GF+N R TLTTG PQ+	
HecA	105	AAAILNEVSPNRSRLAGYLEVAGQAANVVVANPYGITCSGCGFLNTPRLTLTTGTPQFD	164
Orf31ng	215	-AGDFSGFKIRQGNNAVIAGHGLDARDTDF	242
		AG SG +R G+ +I G GLDA +D+	
HecA	165	AAGGLSGLDVRGGDILIDGAGLDASRSY	193

```

25          10          20          30          40          50          60
   orf31-1.pep  MNKTLRYVIFNKRGA VVAEATTKREGKSCADSDSGSAHVKSVPFGTTHAPVCRSNIFS
               |||||
   orf31ng      MNKTLRYVIFNKRGA VVAEATTKREGKSCADSGSGSVYVKSVSFIPTH-----SKAFC
               10          20          30          40          50

30          70          80
   orf31-1.pep  FSLLGFSLCLAVGTANIAFADGI
               ||
   orf31ng      FSALGFSLCLALGTVNIAFADGIITDKAAPKTQQATILQTGNGIPQVNIQTPTSAGVSVN
               60          70          80          90          100          110

```

40

```

      1 ATGAATACTC CTCCTTTTGT CTGTTGGATT TTTTGCAAGG TCATCGACAA
     51 TTTCGGCGAC ATCGGCGTTT CGTGGCGGCT CGCCCGTGTT TTGCACCGCG
    101 AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCGT
    151 GCGCTTTGCC CTGATTGCC  CGATGTTCC  TGCGTTCATC AGGATATTCA
    201 TGTCGCGCAT TGGCATTCCG ATGCGGCAGA TATTGATACC CGC...
```

1 MNTPPFVCWI FCKVIDNEGD IGVSWRLARV LHRELGWQVH LWTDDVSALR  
51 ALCPDLPDVP CVHODIHVRT WHSDAADIDT A..

50

1	ATGAATACTC	CTCCTTTTGT	CTGTTGGATT	TTTTGCAAGG	TCATCGACAA
51	TTTCGGCGAC	ATCGCGGTTT	CGTGGCGGCT	CGCCCGTGTT	TTGCACCGCG
101	AACTCGGTTG	GCAGGTGCAT	TTGTGGACGG	ACGATGTGTC	CGCCTTGCGT

-152-

5  
10  
15  
20

```

151 GCGCTTTGCC CTGATTGCC CGATGTTCCC TGCCTTCATC AGGATATTCA
201 TGTCCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCGCCTGTC
251 CCGATGTCGT CATCGAAACT TTTGCCTGCG ACCTGCCCGA AAATGTGCTG
301 CACATTATCC GCCGACACAA GCCGCTTTGG CTGAATTGGG AATATTTGAG
351 CGCGGAGGAA AGCAATGAAA GGCTGCATCT GATGCCTTCG CCGCAGGAGG
401 GTGTTCAAAA ATATTTTGG TTTATGGGTT TCAGCGAAAA AAGCGGCGGG
451 TTGATACGCG AACGTGATTA CTGCGAAGCC GTCCGTTTCG ATACTGAAGC
501 CCTGCGAGAG CGGCTGATGC TGCCCGAAAA AAACGCCTCC GAATGGCTGC
551 TTTTCGGCTA TCGGAGCGAT GTTTGGGCAA AGTGGCTGGA AATGTGGCGA
601 CAGGCAGGCA GCCCGATGAC ACTGTTGCTG GCGGGGACGC AAATCATCGA
651 CAGCCTCAAA CAAAGCGGCG TTATTCCGCA AGATGCCCTG CAAAACGACG
701 GCGATGTTTT TCAGACGGCA TCCGTCCGCC TCGTCAAAAT CCCTTTCGTG
751 CCGCAACAGG ACTTCGACCA ACTGCTGCAC CTTGCCGACT GCGCCGTCAT
801 CCGCGGCGAA GACAGTTTCG TCGCGGCCCA GCTTGCGGGC AAACCTTTCT
851 TTTGGCACAT CTACCCGCAA GACGAGAATG TCCATCTCGA CAAACTCCAC
901 GCCTTTTGGG ATAAGGCACA CGGTTTCTAC ACGCCCGAAA CCGTGTCCGC
951 ACACCGCCGT CTTTCGACG ACCTCAACGG CGGAGAGGCT TTATCCGCAA
1001 CACAACGCCT CGAATGTTGG CAAACCCTGC AACAACATCA AAACGGCTGG
1051 CCGCAAGGCG CGGAGGATTG GAGCCGTTAT CTTTTCGGGC AGCCGTCAGC
1101 TCCTGAAAAA CTCGTGCCT TTGTTTCAA GCATCAAAA ATACGCTAG

```

This corresponds to the amino acid sequence <SEQ ID 190; ORF32-1>:

25  
30

```

1 MNTPPFVCWI FCKVIDNFGD IGVSWRLARV LHRELGWQVH LWTDDVSALR
51 ALCPDLDPVP CVHQDIHVRT WHSDAADIDT APVPDVVIET FACDLPENVL
101 HIIRRHKPLW LNWEYLSAEE SNERLHLMPS PQEGVQKYFW FMGFSEKSGG
151 LIRERDYCEA VRFDTEALRE RLMLPEKNAS EWLLFGYRSD VWAKWLEMWR
201 QAGSPMTLLL AGTQIIDSLE QSGVIPQDAL QNDGDVFQTA SVRLVKIPFV
251 PQQDFDQLLH LADCAVIRGE DSFVRAQLAG KPFFWHIYPQ DENVHLDKLI
301 AFWDKAHGFY TPETVSAHRR LSDDLNGGEA LSATORLECW QTLQQHQNGW
351 RQGAEDWSRY LFGQPSAPEK LAAVSKHQK IR*w

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF32 shows 93.8% identity over a 81aa overlap with an ORF (ORF32a) from strain A of *N. meningitidis*:

35  
40

```

      10      20      30      40      50      60
orf32.pep MNTPPFVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLDPVP
      |||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
orf32a    MNTPPFSAGXFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLDPVX
      10      20      30      40      50      60

      70      80
orf32.pep CVHQDIHVRTWHSDAADIDTA
      ||||||||||||||||||||
orf32a    CVHQDIHVRTWHSDAADIDTAPXDVVIETFACDLPENVLHIIRRHKPLWLWEYLSAEX
      70      80      90      100     110     120

```

45 The complete length ORF32a nucleotide sequence <SEQ ID 191> is:

50  
55  
60

```

1 ATGAATACTC CTCCTTTTTC TGCTGGANTT TTTTGCAAGG TCATCGACAA
51 TTTTCGCGAC ATCGGCGTTT CGTGGCGGCT TGCCCGTGTT TTGCACCGCG
101 AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCGT
151 GCGCTTTGCC CTGATTGCC CGATGTTTNC TGCCTTCATC AGGATATTCA
201 TGTCCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCGCCTGTC
251 NCGATGTCGT CATCGAAACT TTTGCCTGCG ACCTGCCCGA AAATGTGCTG
301 CACATCATCC GCCGACACAA GCCGCTTTGG CTGAANTGGG AATATTTGAG
351 CGCGGAGGAN AGCAATGAAA GGCTGCACNT GATGCCTTCG CCGCAGGAGA
401 GTGTTCAAAA ATANTTTTGG TTTATGGGTT TCAGCGAANN NAGCGGCGGA
451 CTGATACGCG AACGCGATTA CTGCGAAGCC GTCCGTTTCG ATAGCGGAGC
501 CTTGCGCAAG AGGCTGATGC TTCCCGAAAA AAACGNCCCC GAATGGCTGC
551 TTTTCGGCTA TCGGAGCGAT GTTTGGGCAA AGTGGCTGGA AATGTGGCGA
601 CAGGCAGGCA GTCCGTTGAC ACTTTTGCTG GCNCGGCGCG ANATTATCGA
651 CAGCCTCAAA CAAACGCGCG TTATTCCGCA AGATGCCCTG CAAAACGACG
701 GCGATGTTTT TCAGACGGCA TCCGTCCGCC TCGTCAAAAT CCCTTTCGTG
751 CCGCAACAGG ACTTCGACAA ACTGCTGCAC CTTGCCGACT GCGCCGTCAT

```



5  
 801 CCGCGGCGGAA GACAGTTTCG TGCGCGCCCA GCTTGCGGGC AAACCCCTTCT  
 851 TTTGGCACAT CTACCCGCAA GATGAGAATG TCCATCTCGA CAAACTCCAC  
 901 GCCTTTTGGG ATAAGGCACA CGGTTTCTAC ACGCCCGAAA CCGCATCGGC  
 951 ACACCGCCGC CTTTCAGACG ACCTCAACGG CGGAGAGGCT TTATCCGCAA  
 1001 CACAACGCCCT CGAATGTTGG CAAATCCTGC AACAAACATCA AAACGGCTGG  
 1051 CGGCAAGGCG CGGAGGATTG GAGCCGTTAT CTTTTGGGC AGCCTTCCGC  
 1101 ATCCGAAAAA CTCGCCGCT TGTTCCTCAA GCATCAAAAA ATACGCTAG

This encodes a protein having amino acid sequence <SEQ ID 192>:

10  
 1 MNTPPFSAGX FCKVIDNFGD IGVSWRLARV LHRELGWQVH LWTDDVSALR  
 51 ALCPDLPDVX CVHQDIHVRT WHSDAADIDT APVXDVIET FACDLPENVL  
 101 HIIRRHKPLW LXWEYLSAEX SNERLHXMPS PQESVXKXFW FMGFSEKSGG  
 151 LIRERDYCEA VRFDSGALRK RLMLPEKNXP EWLLFGYRSD VWAKWLEMWR  
 201 QAGSPLTLL AGAXIIDLK QNGVIPQDAL QNDGDVFQTA SVRLVKIPFV  
 251 PQQDFDKLLH LADCAVIRGE DSFVRAQLAG KPFFWHIYPQ DENVHLDKLH  
 15  
 301 AFDKHAHGFY TPETASAHRR LSDDLNGGEA LSATQRLCW QILQQHQNGW  
 351 RQGAEDWSRY LFGQPSASEK LAAFVSKHQK IR\*

ORF32a and ORF32-1 show 93.2% identity in 382 aa overlap:

		10	20	30	40	50	60
20	orf32-1.pep	MNTPPFVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVP					
	orf32a	MNTPPFSAGXFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVX					
		10	20	30	40	50	60
25	orf32-1.pep	70	80	90	100	110	120
		CVHQDIHVRTWHSDAADIDTAPVXDVIETFACDLPENVLHIIRRHKPLWLNWEYLSAEE					
	orf32a	CVHQDIHVRTWHSDAADIDTAPVXDVIETFACDLPENVLHIIRRHKPLWLXWEYLSAEX					
		70	80	90	100	110	120
30	orf32-1.pep	130	140	150	160	170	180
		SNERLHLMPSPOQEGVQKYFWFMGFSEKSGGLIRERDYCEAVRFDTEALRERLMLPEKNAS					
	orf32a	SNERLHXMPSPOQESVXKXFWFMGFSEKSGGLIRERDYCEAVRFDGALRKRLMLPEKNXP					
		130	140	150	160	170	180
35	orf32-1.pep	190	200	210	220	230	240
		EWLLFGYRSDVWAKWLEMWRQAGSPMTLLLAGTQIIDSILKQSGVIPQDALQNDGDVFQTA					
	orf32a	EWLLFGYRSDVWAKWLEMWRQAGSPLTLLLAGAXIIDSILKQNGVIPQDALQNDGDVFQTA					
		190	200	210	220	230	240
40	orf32-1.pep	250	260	270	280	290	300
		SVRLVKIPFVPPQDFDQLHLADCAVIRGEDSFVRAQLAGKPFFWHIYPQDENVHLDKLH					
	orf32a	SVRLVKIPFVPPQDFDKLLHLADCAVIRGEDSFVRAQLAGKPFFWHIYPQDENVHLDKLH					
		250	260	270	280	290	300
45	orf32-1.pep	310	320	330	340	350	360
		AFWDKAHGFYTPETSAHRRLSDDLNGGEALSATQRLCWQTLQQHQNGWRQGAEDWSRY					
	orf32a	AFWDKAHGFYTPETSAHRRLSDDLNGGEALSATQRLCWQILQQHQNGWRQGAEDWSRY					
		310	320	330	340	350	360
50	orf32-1.pep	370	380				
		LFGQPSAPEKLAFAVSKHQKIRX					
	orf32a	LFGQPSASEKLAFAVSKHQKIRX					
		370	380				

# 60 Homology with a predicted ORF from *N.gonorrhoeae*

ORF32 shows 95.1% identity over a 82aa overlap with a predicted ORF (ORF32.ng) from *N.gonorrhoeae*:

orf32.pep MNTPPF-VCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLP 57  
 ||| | |||||

orf32ng MVMNTYAFPVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTTDDVSALRALCPDLP 60  
 orf32.pep DVPCVHQDIHVRTWHSADAADIDTA 81  
 ||| |||||  
 5 orf32ng DVPFVHQDIHVRTWHSADAADIDTAPVPDAVIETFACDLPENVLNIIRRHKPLWLNWEYLS 120

An ORF32ng nucleotide sequence <SEQ ID 193> was predicted to encode a protein having amino acid sequence <SEQ ID 194>:

1 MVMNTYAFPV CWIFCKVIDN FGDIGVSWRL ARVLHRELGW QVHLWTTDDVS  
 51 ALRALCPDLP DVPFVHQDIH VRTWHSADAAD IDTAPVPDAV IETFACDLPE  
 101 NVLNIIRRHK PLWLNWEYLS AESNERLHL MPSPQEGVQK YFWFMGFSEK  
 151 SGGILIRERDY REAVRFDTEA LRRRLVLPEK NAPEWLLFGY RGDVWAKWLD  
 201 MWQQAGSLMT LLLAGAQIID SLKQSGVIPA NALQNEGGVF QTASVRLVKI  
 251 PFVPPQDFDK LLHLADCAVI RGEDSFVRTQ LAGKPFWHI YPDENVHLD  
 301 KLHAFWDKAY GFYTPETASV HRLSDDLNG GEALSATQRL ECGVL\*

15 Further sequencing revealed the following DNA sequence <SEQ ID 195>:

1 ATGAATACAT ACGCTTTTCC TGTCTGTTGG ATTTTGTGCA AGGTCATCGA  
 51 CAATTTTCGGC GACATCGGCG TTTCGTGGCG GCTCGCCCGT GTTTTGCACC  
 101 GCGAAGCTCGG TTGGCAGGTG CATTGTGTGA CGGACGACGT GTCCGCCTTG  
 151 CGCGCGCTTT GTCCCGATTT GCCCGATGTT CCCTTCGTTC ATCAGGATAT  
 201 TCATGTCGCG ACTTGGCATT CCGATGCGGC AGACATGAT ACCGCGCCCG  
 251 TTCCCGATGC CGTTATCGAA ACTTTTGCCT GCGACCTGCC CGAAAATGTG  
 301 CTGAACATCA TCCGCCGACA CAAACCGCTT TGGCTGAATT GGAATATTT  
 351 GAGCGCGGAG GAAAGCAATG AAAGGCTGCA CCTGATGCCT TCGCCGCAGG  
 401 AGGGCGTTCA AAAATATTTT TGGTTTATGG GTTTCAGCGA AAAAAGCGGC  
 25 451 GGGTTGATAC GCGAACGCGA TTACCGCGAA GCCGTCCGTT TCGATACCGA  
 501 AGCCCTGCGC CGGCGGCTGG TGCTGCCCGA AAAAAACGCC CCGAATGGC  
 551 TGCTTTTCGG CTATCGGGGC GATGTTTGGG CAAAGTGGCT GGACATGTGG  
 601 CAACAGGCAG GCAGCCTGAT GACCCTACTG CTGGCGGGGG CGCAAATTAT  
 651 CGACAGCCTC AAACAAAGCG GCGTTATTC GCAAAACGCC CTGCAAAATg  
 30 701 aagcgGTGT CTTTCagacG gcatcgcTcC gccttGTCAA AatcCGTTC  
 751 GTGCcGCAAC AGGAcTTCGA CAAATTGCTG CAcctcgcCG ACTGCGCCGT  
 801 GATACGCGGC GAAGACAGTT TCGTGCGTAC CCAGCTTGCC GGAAAACCTT  
 851 TTTTGTGGCA CATCTACCCG CAAGACGAGA ATGTCCATCT CGACAAACTC  
 901 CAGCCTTTT GGGATAAGGC ATACGGCTTC TACACGCCCG AAACCGCATC  
 35 951 GGTGCACCGC CTCCTTTCCG ACGACCTCAA CGGCGGAGAG GCTTTATCCG  
 1001 CAACACAACG CCTCGAATGT TGGCAAACCC TGCAACAACA TCAAAACGGC  
 1051 TGGCGGCAAG GCGCGGAGGA TTGGAGCCGT TATCTTTTCG GGCAGCCTTC  
 1101 CGCATCCGAA AAACCTGCGC CCTTTGTTTC AAAGCATCAA AAAATACGCT  
 1151 AG

40 This encodes a protein having amino acid sequence <SEQ ID 196; ORF32ng-1>:

1 MNTYAFPVCW IFCKVIDNFG DIGVSWRLAR VLHRELGWQV HLWTTDDVSAL  
 51 RALCPDLPDV PFVHQDIHVR TWHSADAAD IDTAPVPDAVIE TFACDLPENV  
 101 LNIIRRHKPL WLNWEYLSAE ESNERLHLMP SPQEGVQKYF WFMGFSEKSG  
 45 151 GLIRERDYRE AVRFDEALR RRLVLPEKNA PEWLLFGYRG DVWAKWLDWM  
 201 QQAGSLMTLL LAGAQIIDS LKQSGVIPQNA LQNEGGVFQT ASVRLVKIPF  
 251 VPQQDFDKLL HLADCAVIRG EDSFVRTQLA GKPFWHIYP QDENVHLDKL  
 301 HAFWDKAYGF YTPETASVHR LLSDDLNGGE ALSATQRLC WQTLQQHQNG  
 351 WRQGAEDWSR YLFGQPSASE KLAAFVSKHQ KIR\*

ORF32ng-1 and ORF32-1 show 93.5% identity in 383 aa overlap:

50 orf32-1.pep 10 20 30 40 50 59  
 MNTPPF-VCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTTDDVSALRALCPDLPDV  
 ||| |  
 orf32ng-1 MNTYAFPVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTTDDVSALRALCPDLPDV  
 10 20 30 40 50 60  
 55 orf32-1.pep 60 70 80 90 100 110 119  
 PCVHQDIHVRTWHSADAADIDTAPVPDVVIETFACDLPENVLHIIRRHKPLWLNWEYLSAE  
 |  
 orf32ng-1 PFVHQDIHVRTWHSADAADIDTAPVPDAVIETFACDLPENVLNIIRRHKPLWLNWEYLSAE  
 60 70 80 90 100 110 120  
 120 130 140 150 160 170 179

5	orf32-1.pep	ESNERLHLMPSPOEGVQKYFWFMGFSEKSGGLIRERDYCEAVRFDTEALRERLMLPEKNA
	orf32ng-1	ESNERLHLMPSPOEGVQKYFWFMGFSEKSGGLIRERDYREAVRFDTEALRRRLVLPEKNA
10	orf32-1.pep	SEWLLFGYRSDVWAKWLEMWRQAGSPMTLLLAGTQIIDSLSKQSGVIPQDALQNDGDVVFQT
	orf32ng-1	PEWLLFGYRGDVWAKWLDWQQAGSLMTLLLAGAQIIDSLSKQSGVIPQNALQNEGGVFQT
15	orf32-1.pep	ASVRLVKIPFVPPQDFDQLLHLADCAVIRGEDSFVRAQLAGKPFFWHIYPQDENVHLDKL
	orf32ng-1	ASVRLVKIPFVPPQDFDKLLHLADCAVIRGEDSFVRTQLAGKPFFWHIYPQDENVHLDKL
20	orf32-1.pep	HAFWDKANGFYTPETVSAHRRLSDDLNGGEALSATORLECWQTLQHQHONGWRQGAEDWSR
	orf32ng-1	HAFWDKAYGFYTPETASVHRRLSDDLNGGEALSATORLECWQTLQHQHONGWRQGAEDWSR
25	orf32-1.pep	YLFGQPSAPEKLAAFVSKHQKIRX
	orf32ng-1	YLFGQPSASEKLAAFVSKHQKIRX

30 On this basis, including the RGD sequence in the gonococcal protein, characteristic of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF32-1 (42kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 7A shows the results of affinity purification of the His-fusion protein, and Figure 7B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA, giving a positive result. These experiments confirm that ORF32-1 is a surface-exposed protein, and that it is a useful immunogen.

#### Example 24

40 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 197>:

1	..TTGTTCTGCG	GTGTNAAAGT	GGGGCGTTTT	TTCAGCAGTC	CGGCGACGTG
51	GTTTCGGGNC	AAAGACCCTG	TAAATCAGGC	GGTGTTGCGG	CTGTATNCGG
101	ACGAGTGGCG	GCA.ACTTCG	GTACGTTGGA	AAATAGNCGC	AACGTCGCAC
151	AGCCTGTGGC	TCTGCACGCT	GCTCGGAATG	CTGGTGTCGG	TATTGTTGCT
201	GCTTTTGGTG	CGGCAATATA	CGTTCAACTG	GGAAAGCAGC	CTGTTGAGCA
251	ATGCCGCTTC	GGTACGCGCG	GTGAAATGT	TGGCATGGCT	GCCGTCGAAA
301	CTCGGTTTCC	CTGTCCCGCA	TGCGCGGTCTG	GTCATCGAAG	GCCGTCTGAA
351	CGGCAATATT	GCCGATGCGC	GGGCTTGGTC	GGGGCTGCTG	GTCGNCAGTA
401	TCGCCTGCTA	NGGCATCTG	CCGCGCCTG		

50 This corresponds to the amino acid sequence <SEQ ID 198; ORF33>:

1	..LFLRVKVGFR	FSSPATWFRX	KDPVNQAVLR	LYXDEWRXTS	VRWKIXATSH
51	SLWLCTLLGM	LVSVLLLLLV	RQYTFNWEST	LLSNAASVRA	VEMLAWLP SK
101	LGFPVPDARS	VIEGRNLNGNI	ADARAWSGLL	VXSIACXGIL	PRL..

Further work revealed the complete nucleotide sequence <SEQ ID 199>:

```

      1 ATGTTGAATC CATCCGAAA ACTGGTTGAG CTGGTCCGTA TTTTGGACGA
    51 AGGCGGTTTT ATTTTCAGCG GCGATCCCGT ACAGGCGACG GAGGCTTTTC
  101 GCCGCGTGA CCGCAGTACG GAGGAAAAAA TCATCCGTCG GGCGGAGATG
  151 ATTGACAGGA ACCGTATGCT GCGGGAGACG TTGGAACGTG TCGGTGCGGG
    201 GTCGTTCTGG TTGTGGGTGG TGGCGGCGAC GTTTGCATTT TTTACCGGTT
    251 TTTTCAGTAC TTATCTTCTA ATGGACAATC AGGGTCTGAA TTTCTTTTGT
    301 GTTTTGGCGG GCGTGTGGG CATGAATACG CTGATGCTGG CAGTATGGTT
    351 GGCAATGTTG TTCCTGCGTG TGAAAGTGGG GCGTTTTCAG AGCAGTCCGG
  401 CGACGTGGTT TCGGGGCAAA GACCCTGTAA ATCAGGCGGT GTTGCGGCTG
  451 TATGCGGACG AGTGGCGGCA ACCTTCGGTA CGTTGGAAAA TAGGCGCAAC
  501 GTCGCACAGC CTGTGGCTCT GCACGCTGCT CGGAATGCTG GTGTCGGTAT
  551 TGTTGCTGCT TTTGGTGGG CAATATACGT TCAACTGGGA AAGCACGCTG
  601 TTGAGCAATG CCGCTTCGGT ACGCGCGGTG GAAATGTTGG CATGGCTGCC
  651 GTCGAAATC GGTTCCTCTG TCCCGATGCG GCGGGCGGTC ATCGAAGGCC
  701 GTCTGAACGG CAATATTGCC GATGCGCGGG CTTGGTGGG GCTGCTGGTC
  751 GGCAGTATCG CCTGCTACGG CATCCTGCCG CGCCTGCTGG CTGGGTAGT
  801 GTGTAAATC CTTTGAATA CAAGCGAAA CGGATGGAT TTGAAAAGC
  851 CCTATTATCA GCGGCTCATC CGCCGCTGGC AGAACAAAAT CACCGATGCG
  901 GATACGCGTC GGGAAACCGT GTCCGCCGTT TCACCGAAA TCATCTTGAA
  951 CGATGCGCGG AAATGGGCGG TCATGCTGGA GACCGAGTGG CAGGACGCGG
 1001 AATGTTTGA GGCAGGCTG GCGCAGGAAT GGCTGGATA GGGCGTTGCC
 1051 ACCAATCGGG AACAGGTTGC CGCGCTGGAG ACAGAGCTGA AGCAGAAACC
 1101 GGCACAATG CTTATCGGCG TGCGCGCCCA AACTGTGCCG GACCGCGGCG
 1151 TGTTGCGGCA GATTGTCCGA CTCTCGGAAG CGGCGCAGGG CGGCGCGGTG
 1201 GTGCAGCTTT TGGCGGAACA GGGGCTTTCA GACGACCTTT CGGAAAAGCT
 1251 GGAACATTGG CGTAACGCGC TGGCCGAATG CGGCGCGGCG TGGCTTGAGC
 1301 CTGACAGGGC GGCGCAGGAA GGGCGTTTGA AAGACCAATA A

```

This corresponds to the amino acid sequence <SEQ ID 200; ORF33-1>:

```

  30      1 MLNPSRKLVE LVRILDEGGF IFSGDPVQAT EALRRVDGST EEKIIRRAEM
    51 IDNRMLRET LERVAGSFW LWVVAATFAE FTGFSVTYLL MDNQGLNFFL
  101 VLAGVLMNT LMLAVWLAML FLRVKVGREF SSPATWFRGK DPNQAVLRL
  151 YADEWRQPSV RWKIGATSHS LWLCTLLGML VSVLLLLLVR QYTFNWESTL
  201 LSNAASVRAV EMLAWLPSKL GFPVPDARAV IEGRLNGNIA DARAWSGLLV
  251 GSIACYGILP RLLAWVVCKI LLKTSSEGLD LEKPYQAVI RRWQNKITDA
  301 DTRRETSAV SPKILNDAP KWAVMLETEW QDGEWFEGRL AQEWLDKQVA
  351 TNREQVALE TELQKPAQL LIGVRAQTPV DRGVLRQIVR LSEAAQQGAV
  401 VQLLAQGLS DDLSEKLEHW RNALAECEGAA WLEPDRAAQE GRLKDQ*

```

Computer analysis of this amino acid sequence gave the following results:

#### 40 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF33 shows 90.9% identity over a 143aa overlap with an ORF (ORF33a) from strain A of *N. meningitidis*:

```

  45      orf33.pep                                10    20    30
      LFLRVKVGREFSSPATWFRXKDPVNQAVLR
      orf33a      LMDNQGLNFFLVLAGVXGMNTLMLAVWLAMLFLRVKVGREFSSPATWFRGKDPVNQAVLR
      90      100      110      120      130      140

  50      orf33.pep                                40    50    60    70    80    90
      LYXDEWRXTSVRWKIXATSHSLWLCTLLGMLVSVLLLLLVRQYTFNWESTLLSNAASVRA
      orf33a      LYADEWRXPSVRWKIGATSHSLWLCTLLGMLVSVLLLLLVRQYTFNWESTLLGDSSSVRL
      150      160      170      180      190      200

  55      orf33.pep                                100    110    120    130    140
      VEMLAWLPSKLGFPVPDARSVIEGRLNGNIA DARAWSGLLVXSIACXGILPRL
      orf33a      VEMLAWLPAKLGFVPDARAVIEGRLNGNIA DARAWSGLLVGSIACYGILPRLAWAVCK
      210      220      230      240      250      260

  60      orf33a      ILXTSSEGLDLEKXXXXXXIRRWQNKITDADTRRETSAVSPKIVLNDAPKWAVMLETE
      270      280      290      300      310      320

```

The complete length ORF33a nucleotide sequence <SEQ ID 201> is:

```

1  ATGTTGAATC CATCCCGAAA ACTGGTTGAG CTGGTCCGTA TTTTGAAGA
51 AGGCGGCTTT ATTTTCAGCG GCGATCCCGT GCAGGCGACG GAGGCTTTGC
101 GCCGCGTGA CCGCAGTACG GAGGAAAAAA TCATCCGTCG GCGGAAGATG
5  151 ATCGACAGGA ACCGTATGCT CCGGGAGACG TTGGAACGTG TCGGTGCGGG
201 GTCGTTCTGG TTGTGGGTGG CCGCGGCGAC GTTTGCGTTT NTTACCGNTT
251 TTTCAGTTAC TTATCTTCTA ATGGACAATC AGGGTCTGAA TTTCTTTTGT
301 GTTTTGGCGG GCGTGNTGGG CATGAATACG CTGATGCTGG CAGTATGGTT
10 351 GGCAATGTTG TTCCTGCGCG TGAAAGTGGG GCGTTTTCAG AGCAGTCCGG
401 CGACGTGGTT TCGGGGCAAA GACCCTGTCA ATCAGGCGGT GTTGGCGCTG
451 TATGCGGACG AGTGGCGGCN ACCTTCGGTA CGTTGGAATA TAGGCGCAAC
501 GTCGCACAGC CTGTGCTCTT GCACGCTGCT CGGAATGCTG GTGTCGGTAT
551 TGTGCTGCTG TTTGGTGGCG CAATATACGT TCAACTGGGA AAGCAGCTG
601 TTGGGCGATT CGTCTTCGGT ACGGCTGGTG GAAATGTTGG CATGGCTGCC
15 651 TGCGAAACTG GGTTCCTCCG TGCCTGATGC GCGGGCGGTC ATCGAAGGTC
701 GATACGACGG CAATATTGCC GATGCGCGGG CTTGGTGGGG GCTGCTGGTC
751 GGCAGTATCG CCTGCTACGG CATCCTGCCG CGCCTCTTGG CTTGGGCGGT
801 ATGCAAAATC CTTNTGNAAA CAAGCGAAAA CGGCTGGGAT TTGAAAAGC
851 NCNNNNNTCN NNCNGTCATC CGCCGCTGGC AGAACAAAAT CACCGATGCG
20 901 GATACGCGTC GGGAAACCGT GTCCGCGGTT TCGCGGAAA TCGTCTTGAA
951 CGATGCGCCG AAATGGGCGG TCATGCTGGA GACCGAATGG CAGGACGGCG
1001 AATGGTTTCA GGGCAGGCTG GCGCAGGAAT GGCTGGATAA GGGCGTTGCC
1051 GCCAATCGGG AACAGGTTGC CGCGCTGGAG ACAGAGCTGA AGCAGAAACC
1101 GCGCAACTG CTTATCGGCG TGCGCGCCCA AACTGTGCCC GACCGCGGCG
25 1151 TGTGCGGCA GATCTCCGGA CTTTCGGGAG CGGCGCAGGG CGGCGCGGTG
1201 GTGCANCTTT TGGCGGAACA GGGGCTTTCA GACGACCTTT CGGAAAAGCT
1251 GGAACATTGG CGTAACGCGC TGACCGAATG CGGCGCGGCG TGGCTGGAAC
1301 CCGACAGAGC GCGCAGGAA GGCCGTCTGA AAACCAACGA CCGCACTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 202>:

```

30 1  MLNPSRKLVE LVRILEEGGF IFSGDPVQAT EALRRVDGST EEKIIRRAKM
51  IDNRNMLRET LERVAGSFW LWVAAATFAF XTXFSVTYLL MDNQGLNFFL
101 VLAGVXGMNT LMLAVWLAML FLRVKVGREF SSPATWFRGK DPVNQAVLRL
151 YADEWRXPSV RWKIGATSHS LWLCTLLGML VSVLLLLLVR QYTFNWESTL
35 201 LGDSSSVRLV EMLAWLPAKL GFPVPDARAV IEGRNLGNIA DARAWSGLLV
251 GSIACYGILP RLLAWAVCKI LXXTSENGLD LEKXXXXXXI RRWQNKITDA
301 DTRRETSAV SPKIVLNDAP KWAVMLETEW QDGEWFEGRL AQEWLDKGVA
351 ANREQVALE TELKQKPAQL LIGVRAQTPV DRGVLRQIVR LSEAAQGGAV
401 VXLLAEQGLS DDLSEKLEHW RNALTECGAA WLEPDRAAQE GRLKTNDR*

```

ORF33a and ORF33-1 show 94.1% identity in 444 aa overlap:

```

40      10      20      30      40      50      60
orf33a.pep  MLNPSRKLVELVRILEEGGFIFSGDPVQATEALRRVDGSTEEKIIRRAKMIDNRNMLRET
          |||
orf33-1     MLNPSRKLVELVRILDEGGFIFSGDPVQATEALRRVDGSTEEKIIRRAEMIDNRNMLRET
          |||

45      10      20      30      40      50      60
orf33a.pep  LERVAGSFWLWVAAATFAFXTXFSVTYLLMDNQGLNFFLVLAGVXGMNTLMLAVWLAML
          |||
orf33-1     LERVAGSFWLWVVAATFAFFTGFSTYLLMDNQGLNFFLVLAGVXGMNTLMLAVWLAML
          |||

50      70      80      90      100     110     120
orf33a.pep  LERVAGSFWLWVAAATFAFXTXFSVTYLLMDNQGLNFFLVLAGVXGMNTLMLAVWLAML
          |||
orf33-1     LERVAGSFWLWVVAATFAFFTGFSTYLLMDNQGLNFFLVLAGVXGMNTLMLAVWLAML
          |||

55      130     140     150     160     170     180
orf33a.pep  FLRVKVGREFSSPATWFRGKDPVNQAVLRLYADEWRXPSVRWKIGATSHSLWLCTLLGML
          |||
orf33-1     FLRVKVGREFSSPATWFRGKDPVNQAVLRLYADEWRQPSVRWKIGATSHSLWLCTLLGML
          |||

60      190     200     210     220     230     240
orf33a.pep  VSVLLLLLVRQYTFNWESTLLGDSSSVRLVEMLAWLPAKLGFPVPDARAVIEGRNLGNIA
          |||
orf33-1     VSVLLLLLVRQYTFNWESTLLSNAASVRVEMLAWLPSKLGFPVPDARAVIEGRNLGNIA
          |||

65      250     260     270     280     290     300
orf33a.pep  DARAWSGLLVGSIACYGILPRLAWAVCKILXXTSENGLDLEKXXXXXXIRRWQNKITDA
          |||

```

	orf33-1	DARAWSGLLVGSIACYGILPRLLAWVVCKILLKTSENGLDLEKPYQAVIRRWQNKITDA	250	260	270	280	290	300
			310	320	330	340	350	360
5	orf33a.pep	DTRRETSAVSPKIVLNDAPKWAVMLETEWQDGEWFEGRLAQEWLDKGVANREQVAALE						
	orf33-1	DTRRETSAVSPKIIILNDAPKWAVMLETEWQDGEWFEGRLAQEWLDKGVATNREQVAALE						
10	orf33a.pep	TELKQKPAQLLIGVRAQTVPDRGVLRQIVRLSEAAQGGAVVXLLAEQGLSDDLSEKLEHW	370	380	390	400	410	420
	orf33-1	TELKQKPAQLLIGVRAQTVPDRGVLRQIVRLSEAAQGGAVVQLLAEQGLSDDLSEKLEHW						
15	orf33a.pep	RNALTECGAAWLEPDRAAQEGRLKTNDRTX	430	440	450			
	orf33-1	RNALAECGAAWLEPDRAAQEGRLKDQX						
20			430	440				

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF33 shows 91.6% identity over a 143aa overlap with a predicted ORF (ORF33.ng) from *N. gonorrhoeae*:

25	orf33.pep	LFLRVKVGRRFFSSPATWFRXKDPVNQAVLR	30
	orf33ng	LMDNQGLNFFLVLAGVLMNTLM LAVWLATLFLRVKVGRRFFSSPATWFRGKGPVNQAVLR	100
30	orf33.pep	LYXDEWRXTSVRWKIXATSHSLWLCTLLGMLVSVLLLLLVLRQYTFNWESTLLSNAASVRA	90
	orf33ng	LYADQWRQPSVRWKIGATAHSLWLCTLLGMLVSVLLLLLVLRQYTFNWESTLLSNAASVRA	160
35	orf33.pep	VEMLAWLPSKLGFPVPDARSVIEGRNLNGNIADARAWSGLLVXSIACXGILPRL	143
	orf33ng	VEMLAWLPSKLGFPVPDARAVIEGRNLNGNIADARAWSGLLVGSIVCYGILPRLLAWVVCK	220

An ORF33ng nucleotide sequence <SEQ ID 203> was predicted to encode a protein having amino acid sequence <SEQ ID 204>:

40	1	MIDRDRMLRD	TLERVAGSF	WLWVVASMM	FTAGFSGYL	LMDNQGLNFF
	51	LVLAVGLGMN	TLMLAVWLAT	LFLRVKVGRR	FSSPATWFRG	KGPVNQAVLR
	101	LYADQWRQPS	VRWKIGATAH	SLWLCTLLGM	LVSLLLLLV	RQYTFNWEST
	151	LLSNAASVRA	VEMLAWLPSK	LGFPVPDARA	VIEGRNLNGNI	ADARAWSGLL
	201	VGSIVCYGIL	PRLLAWVVCK	ILLKTSENGL	DLEKTYQAV	IRRWQNKITD
	251	ADTRRETUSA	VSPKIVLND	PKWALMLETE	WQDQWFEGR	LAQEWLDKGV
45	301	AANREQVAAL	ETELKQKPAQ	LLIGVRAQTV	PDRGVLRQIV	RLSEAAQGGA
	351	VVQLLAEQGL	SDDLSEKLEH	WRNALTECGA	AWLEPDRVAQ	EGRLKDQ*

Further sequence analysis revealed the following DNA sequence <SEQ ID 205>:

50	1	ATGTTGaAtC	CATCCCGaAA	ACTGgttgag	ctGgTCCgtA	Ttttgaataa
	51	aggggggtTTT	attttcagcg	gcgatcctgt	gcaggcgacg	gaggctttgc
	101	gccgcgtgga	cggcAGTACG	GAGgAaaaaa	tcttcgctcg	GGCGGAGAtg
	151	atcgACAGGg	accgtatgtt	gcgggACaCg	TtggaaCGTG	TGCGTGCGgg
	201	gtcgtTctgG	TTATGGGTGG	TggtggCatC	gATGATGTtt	aCCGCCGgAT
	251	TTTCAGgcac	ttatCttCTG	ATGGACaatC	AGGGGCTgAA	TtTCTTTTAA
	301	GTTTtggcgG	GAGTGTtggG	CATGaatacG	ctgATGCTGG	CAGTATGGtt
	351	gGCAACGTTG	TTCCTGCGCG	TGAAAGTGGG	ACGGTTTTC	AGCAGTCCGG
55	401	CGACGTGGT	TCGGGGCAAA	GGCCTGTAA	ATCAGGCGGT	GTTCGGGCTG
	451	TATGCGGACC	AGTGGCGGCA	ACCTTCGGTA	CGATGGAAAA	TAGGCGCAAC
	501	GCGCACAGC	TTGTGGCTCT	GCACGCTGCT	CGGAATGCTG	GTGTGGGTAT
	551	TGCTGCTGCT	TTTGGTGGCG	CAATATACGT	TCAACTGGGA	AAGCACGCTG
	601	TTGAGCAATG	CCGCTTCGGT	ACGCGCGGTG	GAAATGTTGG	CATGGCTGCC
	651	GTCGAACTC	GTTTCCCTG	TCCCGATGC	GCGGGCGGTC	ATCGAAGGTC
60	701	GTCTGAACGG	CAATATTGCC	GATGCGCGGG	CTTGGTGGG	GCTGCTGGTC
	751	GGCAGTATCG	TCTGCTACGG	CATCCTGCCG	CGCCTCTTGG	CTTGGGTAGT

801 GTGTAAATC CTTTGA AAA CAAGCGAAAA CGGAttgGAT TTGGA AAAAA  
 851 CCTATTATCA GGCGGTCATC CGCCGCTGGC AGAACAAAAT CACCGATGCG  
 901 GATACGCGTC GGGAAACCGT GTCCGCCGTT TCGCcgAAAA TCGTCTTGAA  
 951 CGATGCGCCG AAATGGGCGC TCATGCTGGA GACCGAGTGG CAGGACGGCC  
 5 1001 AATGGTTCGA GGGCAGGCTG GCGCAGGAAT GGCTGGATAA GGGCGTTGCC  
 1051 GCCAATCGGG AACAGGTTGC CGCGCTGGAG ACAGAGCTGA AGCAGAAACC  
 1101 GGCGCAACTG CTTATCGGCG TACGCGCCCA AACTGTGCCG GACCGGGGCG  
 1151 TGCTGCGGCA GATTGTGCGG CTTTCGGAAG CGGCGCAGGG CGGCGCGGTG  
 1201 GTGCAGCTTT TGGCGGAACA GGGGCTTCA GACGACCTTT CGGAAAAGCT  
 10 1251 GGAACATGG CGTAACGCGC TGACCGAATG CGGCGCGGCG TGGCTTGAGC  
 1301 CTGACAGGGT GGCAGGAA GGCCGTTGA AAGACCAATA A

This encodes a protein having amino acid sequence <SEQ ID 206; ORF33ng-1>:

1 MLNPSRK LVE LVRILNKG GF IFSGDPVQAT EALRRVDGST EEKIFRAEM  
 51 IDRDRMLRDT LERVAGSEFW LWVVVASMFF TAGFSGTYLL MDNQGLNFFL  
 15 101 VLAGVLMNT LMLAVLATL FLRVKVG RFF SSPATWFRGK GPVNQAVLRL  
 151 YADQWRQPSV RWKIGATAHS LWLCTLLGML VSVLLLLLLVR QYTFNWESTL  
 201 LSNAASVRV EMLAWLPSKL GFPVPDARV IEGRNLGNIA DARAWSGLLV  
 251 GSIVCYGILP RLLAWVVCKI LLKTSENGLD LEKTYQAVI RRWQNKITDA  
 301 DTRRET VSAV SPKIVLNDAP KWALMLET EW QDGQWFEGR L AQEWLDKGVA  
 20 351 ANREQVALE TELKQKPAQL LIGVRAQT VP DRGVLRQIVR LSEAAQGGAV  
 401 VQLLAEQGLS DDLSEKLEHW RNALTECGAA WLEPDRVAQE GRLKDQ\*

ORF33ng-1 and ORF33-1 show 94.6% identity in 446 aa overlap:

		10	20	30	40	50	60
25	orf33-1.pep	MLNPSRK	LVRLDE	GGFIF	SGDPVQ	ATEALR	VDGST
	orf33ng-1	MLNPSRK	LVRLDE	GGFIF	SGDPVQ	ATEALR	VDGST
		10	20	30	40	50	60
30	orf33-1.pep	LERVRAG	SFWLW	VVAAT	FAFFT	GFVS	TYLLMD
	orf33ng-1	LERVRAG	SFWLW	VVAAT	FAFFT	GFVS	TYLLMD
		70	80	90	100	110	120
35	orf33-1.pep	FLRVKVG	RFFSS	PATWFR	GKDPV	NQAVLR	LYADE
	orf33ng-1	FLRVKVG	RFFSS	PATWFR	GKDPV	NQAVLR	LYADE
		130	140	150	160	170	180
40	orf33-1.pep	VSVLLLL	LVRYT	FNWES	TLLS	NAASV	RAVEM
	orf33ng-1	VSVLLLL	LVRYT	FNWES	TLLS	NAASV	RAVEM
		190	200	210	220	230	240
45	orf33-1.pep	DARAWSG	LLVGS	IACYG	ILPR	LLAWV	VCKIL
	orf33ng-1	DARAWSG	LLVGS	IACYG	ILPR	LLAWV	VCKIL
		250	260	270	280	290	300
50	orf33-1.pep	DTRRET	VSAV	SPKIL	NDAP	KWAV	MLET
	orf33ng-1	DTRRET	VSAV	SPKIL	NDAP	KWAV	MLET
		310	320	330	340	350	360
55	orf33-1.pep	TELKQK	PAQL	LIGV	RAQT	VPDR	GVLR
	orf33ng-1	TELKQK	PAQL	LIGV	RAQT	VPDR	GVLR
		370	380	390	400	410	420
60	orf33-1.pep	RNALAE	CGAA	WLEP	DRAA	QEGR	LKDQ
	orf33ng-1	RNALAE	CGAA	WLEP	DRAA	QEGR	LKDQ
		430	440				
65	orf33-1.pep						
	orf33ng-1						

orf33ng-1

 RNALTECGAAWLEPDRVAQEGRLKDQX  
 430 440

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is  
 5 predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be  
 useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 25**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 207>:

```

10      1  ..CAGAAGAGTT TGTCGAGAAT TTCTTTATGG GGTTTGGGCG GCGTGTTTTT
      51  CCGGGTGTCC GGTCTGGTAT GGTTTCTTTT GGGCGTTTCT TT.GAGTGCG
     101  CCTGTTTTC GGGTGTTCCT TTTCGGGGTT CGGGACGGGG GACGTTTGTG
     151  GGCAGTACGG GGGTTTCTTT GAGTGTGTTT TCAGCTTGTG TTCC.GGCGT
     201  CGTCCGCTG CCTGTGCGTT TGAGCTGTGT CGGCAGGTG CG..GTTTGA
     251  CCCGTTTTC CTTGGGTGCG GCAGGGGACG TCATTCTCCT GCCGCTTCG
     15  301  TCTGTGCCGT CCGCTGTGC GGGTTCGGAT GAGGCGGCGT GGTGGTGTTC
     351  GGGTTGGGCG GCATCTTGT CCGACTACGC CGTTTGGCAG CCAGAATTCTG
     401  GTTTCGCGGG GGCTGTCGGT GTGTTGCGGT TCGGCTTGAA GGGTTTGTG
     451  GTCC..
  
```

This corresponds to the amino acid sequence <SEQ ID 208; ORF34>:

```

20      1  ..QKSLSRISLW GLGGVFFGVG GLVWFSLGVS XECACFSGVG FRGSGRGTFV
     51  GSTGVSLVVF SACVXGVVRL PVGLSCVGR LXXLTRFFLGA AGDVILLPLS
     101  SVPSGCAGSD EAAWWC SGWA ASCPTT PFGS QNSVSRGLSV CCGSA*RVLS
     151  S..
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 209>:

```

25      1  ATGATGATGC CGTTCATAAT GCTTCCTTGG ATTGCKGGTG TGCCTGCCGT
     51  GCCGGGTCAG AATAGGTTGT CCAGAATTTC TTTATGGGGT TTGGGCGGGC
     101  TGTTTTTCGG GGTGTCCGGT TTGGTATGGT TTTCTTTGGG CGTTTCTTTG
     151  GGCTGCGCCT GTTTTTCCGG TGTTCTTTT CGGGGTTCCG GACGGGGGAC
     201  GTTTGTGGGC AGTACGGGGG TTTCTTTGAG TGTGTTTTCA GCTTGTGTTT
     30  251  CGGCGTCGTC CGGCTGCCGT TCGGTTTGAG CTGTGTCCGC AGGTGCGGCT
     301  TTGACCCGGT TTTTCTTGGG TGCGGCAGGG GACGGCAGTC CGGTGCCGCT
     351  TTCGTCTGTG CCGTCCGGCT GTGCGGGTTC GGATGAGGCG GCGTGGTGGT
     401  GTTCGGGTTG GCGGGCATCT TGTCCGACTA CGCCGTTTGG CAGCCAGAAT
     451  TCGGTTTCGC GGGGGCTGTC GGTGTGTTGC GGTTGCGCTT GAAGGGTTTTT
     35  501  GTCGCCGTTT GGGTTGAATG TGCTGACGAT GCCTATTGCC AATGCGCCGA
     551  TGGCGGCGAT ACAGATGAGC AATACGGCGC GTATCAGGAG TTTGGGGGTC
     601  AGCCTGAAGG GTTTGTTCGG TTTTCTTGGC ATTTTGATTG TGCTTTTGGG
     651  GTGTCGGGCA ATGCCGTCTG AAGGCGGTTC AGACGGCATT GCCGAGTCAG
     701  CGTTGGACGT AGTTTGGTA GAGGGTGATG ACTTTTGTGA CGCCGACGGT
     40  751  GGTGCTGACT TTTTGGGTAA TCTGCGCCTG TTCTTCGGGG GTGAGGATGC
     801  CCATAACGTA GGTACGTTG CCGTAGGTAA CGATTTTGAC GCGCGCCTGT
     851  TTGGCGGGGC TGATGCCCAA CAGCGTGGCG CGGACTTTGG ATGTGTTCCA
     901  AGTGTCGCCG GCGATGTCGC CGGCAGTGCG CGGCAGGGAG GCGACGTAA
     951  TATAGTTGTA CACGCCTTCG GCGGCCTGTT CGGAACGTGC AATCTGACCG
     45  1001  ACGAACTGTT TTTCCGCTTC GGTGGCGACT TGTCCGAGCA GCAGCAGGTG
     1051  CCGGTTGTAG CCGACGACGG AGATTGCGG CGTGTAGCCT TTGGTTTGGT
     1101  TGTTTTGGCG CAGATAGGAA CGGGCGGTGG TTTGATACG CAACGCCATA
     1151  ACGTTGTCGT CGGTTTGGCG GCCGGTGGTT CGGCGGTCTG CGGCGGATT
     1201  CGCGCCGACG GCGGCGCTTC CGATTACTGC GCTGACGCAG CCGCTAAGGG
     1251  CAAGGCTGAA AATGGCGGCA ATCAGGGTGC GGACGCTGTG CGGTTTGGGT
     1301  TTCATCGGGT GCTTCCTTTC TTGGGCGTTT CAGACGGCAT TGCTTTGCGC
     1351  CATGCCGTCT GA
  
```

This corresponds to the amino acid sequence <SEQ ID 210; ORF34-1>:

```

55      1  MMMPFIMLPW IAGVPAVPGQ NLSRISLWG LGGVFFGVSG LVWFSLGVSL
     51  GCACFSGVSF RSGRGTFVG STGVSLVFS ACVPASSGCL SV*AVSAGCG
     101  LTRFFLGAAG DGSPLPLSSV PSGCAGSDEA AWWCSGWAAS CPTT PFGSQN
     151  SVSRGLSVCC GSA*RVLS PF GLNVLTMPIA NAPMAAIQMS NTARIRSLGV
  
```



-161-

201 SLKGLFGFFA ILIVLLGCRA MPSEGGSDGI AESALDVVLV EGDDFLYADG  
 251 GADFLGNLRL FFGGEDAHNV GYVAVGNDFD ARLCGGADAQ QRGADFGCVF  
 301 SVAGDVAGSA RQGGDGNIVV HAFGGLFGTC NLTDLEFFAF GDDLSEQQQV  
 351 AVVADDGDLG RVAFLVLVLA QIGTGGGFDT QRHNVVVGLR AGGSAVDGGF  
 401 RADGGADYDC ADAAAKGKAE NGGNQAGDV RFGFHRVLPF LGVSDGIALR  
 451 HAV\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF34 shows 73.3% identity over a 161aa overlap with an ORF (ORF34a) from strain A of *N.*

*meningitidis*:

		10	20	30	
orf34.pep		QKSLSRISLWGLGGVFFGVSGLVWVSLGVSXE-----CAC			
orf34a	MMXPXIMLPWIAGVPAVPGQKRLSRXSLWGLGGXFFGVSGLVWVSLGVSXSLGVSXGCAC				
		10	20	30	40
		40	50	60	70
orf34.pep	FSGVSFRGSGRGTFTVGSTGVLSLVFSACVXGVVRLPVGLSCVGRLLX-----LTRFFLGA				
orf34a	FSGVSFRGSGRGTFTVGSTGVLSLVFSACA-----PASSGCLSVXAVSAGCGLTRXFXGA				
		70	80	90	100
		100	110	120	130
orf34.pep	AGDVILLPLSSVPSGCAGSDEAAWVCSGWAASCPTTFPGSQNSVSRGLSVCCGSAXRVLS				
orf34a	AGDGSPLPLSSVPSGCAGADEEAXXCSGWAASCPTTFPGSQNSVSRGLSVCCGSVWRVLS				
		120	130	140	150
		160	170		
orf34.pep	S				
orf34a	PFGXNVLTMPIANAPMAVIQMSNTARIRSLGVSLKGLFXFFAILIVLLGCRAMPSEGGSD				
		180	190	200	210
		220	230		

The complete length ORF34a nucleotide sequence <SEQ ID 211> is:

1 ATGATGATNC CGTTNATAAT GCTTCCTTGG ATTGCGGGTG TGCCTGCCGT  
 51 GCCGGGTGAG AAGAGGTTGT CGAGAANTTC TTTATGGGGT TAGGCGGCN  
 101 TGTTTTTCGG GGTGTCCGGT TTGGTATGGT TTTCTTTGGG CGTTTCTNNT  
 151 TCTTTGGGTG TTTCTNTGGG CTGTGCCTGT TTTTCGGGTG TTTCTTTTCG  
 201 GGGTTCGGGA CGGGGGACGT TTGTGGGCAG TACNGGGGT TCTTTGAGTG  
 251 TGTTTTTCAGC TTGTGCTCCG GCGTCGTCCG GCTGCCTGTC GGTTTNAGCT  
 301 GTGTCCGGCAG GTTCCGGTTT GACCCGGNTT TTCTTNGGTG CGGCAGGGGA  
 351 CGGCAGTCCG CTGCCGCTTT CGTCTGTGCC GTCCGGCTGT GCGGGTGCAG  
 401 ATGAGGAGGC GTNGTNGTGT TCGGGTTGGG CGGCATCTTG TCCGACTACG  
 451 CCGTTTGCA GCCAGAATTC GGTTCGCGG GGGCTGTCGG TGTGTTGCGG  
 501 TCCGGTNTGG AGGGTTTGT CNCCGTTCCG GTNGAATGTG CTGACGATGC  
 551 CTATTGCCAA TCGCCGATG GCGGTGATAC AGATGAGCAA TACGGCGCGT  
 601 ATCAGGAGTT TGGGGGTGAG CCTGAAGGGT TTGTTCTGTT TTTTGGCCAT  
 651 TTTGATTGTG CTTTGGGGT GTCCGGCAAT GCCGTCTGAA GGCGGTTCAG  
 701 ACGGCATTGC CGAGTCAGCG TTGGACGTAG TTTNGGTAGA GGGTGATGAC  
 751 TTTTGTACG CCGACGGTGG TGCTGACTTT TTGGGTAATC TGCGCCTGTT  
 801 CTTCCGGGGT GAGGATGCCC ATAACGTAGG TTACGTGTCG GTAGGTAACG  
 851 ATTTTGACGC GCGCCTGTGT GCGGGGGCTG ATGCCCAACA GCGTGCGCG  
 901 GACTTTGGAT GTGTTCCAAG TGTCGCCGCG GATGTCGCCG GCAGTGCGCG  
 951 GCAGGGAGGC GACGGTAATG TANTTGTACA CGCCTTCGGC GGCCTGTTTC  
 1001 GAACGTGCAA TCTGACCGAC GAACTGTTTC TCGCCTTCGG TGGCGACTTG  
 1051 TCCGAGCAGC AGCAGGTGGC GGTGTAGACC GACAACGGAG ATTTGGGGCG  
 1101 TGTANCTTT GGTTTGGTTG TTTTGGCGCA GATAGGAGCG GGCCTGTTTC  
 1151 TCGATACGCA GCGCCATTAC GTTGTGTCG GTTNGCGCGC CGGTGTTTCG  
 1201 GCGGTCGACG GCGGATTTTC GCGGACCGC CGCGCCGCGC ACGACTGCGC  
 1251 TCAGCGAGCC GCGGAGGGA AGGCTGAGGA CGGCGGCAGT CAGGGTGCAG  
 1301 ACGGTGTGCG GTTTGGGTTT CATCGGGTGC TTCCTTTCTT GGGCGTTTCA  
 1351 GACGGCATTG CTTTGCAGCA TGCCGTCTGA

This encodes a protein having amino acid sequence <SEQ ID 212>:

```

      1  MMXPXIMLPW IAGVPAVPGQ KRLSRXSLWG LGGXFFGVSG LVWFSLGVSX
    51  SLGV SXGCAC FSGVSFRGSG RGT FVGSTGV SLSVFSACAP ASSGCLSVXA
   101  VSAGCGLTRX FXGAAGDGSP LPLSSVPSGC AGADEEAXXC SGWAASCPTT
   151  PFGSQNSVSR GLSVCCGSVW RVLSPFGXNV LTMPIANAPM AVIQMSNTAR
   201  IRLGVS LKGLFXFFAILIV LLGCRAMPSE GSGDGLAESL LDVVXVEGDD
   251  FLYADGGADF LGNLR LFFGG EDHNVGYVA VGNDFDARLC GGADAQQRGA
   301  DFGCVPSVAG DVAGSARQGG DGNVXVHAFG GLFGTCNLTD ELFLAFGGDL
   351  SEQQQVAVVA DNGDLGRVXF GLVVL AQIGA GGGFDTQRHY VVVGXRAGGS
   401  AVDGGFRADR RAADDCA DAA AEGKAEDGGS QGADGVRFGF HRVLPFLGVS
   451  DGIALRHAV*

```

ORF34a and ORF34-1 show 91.3% identity in 459 aa overlap:

```

      10      20      30      40      50      60
or34a.pep  MMXPXIMLPW IAGVPAVPGQ KRLSRXSLWG LGGXFFGVSG LVWFSLGVSX SLGV SXGCAC
15 or34-1    MMMPFIMLPW IAGVPAVPGQ NLRSLRISLWGLGGVFFGVSG LVWFSLGVS L-----GCAC
      10      20      30      40      50
      70      80      90     100     110     120
20 or34a.pep  FSGVSFRGSG RGT FVGSTGV SLSVFSACAP ASSGCLSVX AVSAGCGLTRX FXGAAGDGSP
   or34-1    FSGVSFRGSG RGT FVGSTGV SLSVFSACAP ASSGCLSVX AVSAGCGLTR FFLGAAGDGSP
      60      70      80      90     100     110
      130     140     150     160     170     180
25 or34a.pep  LPLSSVPSGC AGADEEAXXC SGWAASCPTT PFGSQNSVSR GLSVCCGSV VRVLS PFGXNV
   or34-1    LPLSSVPSGC AGSDEAAW C SGWAASCPTT PFGSQNSVSR GLSVCCGS SXRVLSP FGLNV
      120     130     140     150     160     170
      190     200     210     220     230     240
30 or34a.pep  LTMPIANAPM AVIQMSNTAR IRLGVS LKGLFXFFAILIV LLGCRAMPSE GSGDGLAESL
   or34-1    LTMPIANAPM AAIQMSNTAR IRLGVS LKGLFGFFAILIV LLGCRAMPSE GSGDGLAESL
      180     190     200     210     220     230
      250     260     270     280     290     300
40 or34a.pep  LDVVXVEGDD FLYADGGADF LGNLR LFFGG EDHNVGYVA VGNDFDARLC GGADAQQRGA
   or34-1    LDVVLEVGDD FLYADGGADF LGNLR LFFGG EDHNVGYVA VGNDFDARLC GGADAQQRGA
      240     250     260     270     280     290
      310     320     330     340     350     360
45 or34a.pep  DFGCVPSVAG DVAGSARQGG DGNVXVHAFG GLFGTCNLTD ELFLAFGGDL SEQQQVAVVA
   or34-1    DFGCVPSVAG DVAGSARQGG DGNVXVHAFG GLFGTCNLTD ELFFAFGGDL SEQQQVAVVA
      300     310     320     330     340     350
      370     380     390     400     410     420
50 or34a.pep  DNGDLGRVXF GLVVL AQIGA GGGFDTQRHY VVVGXRAGGS AVDGGFRADRR AADDCA DAA
   or34-1    DDGDLGRVAF GLVVL AQIGTGGG FDTQRHNVVVG LRAGGS AVDGGFRADGG ASDYCA DAA
      360     370     380     390     400     410
      430     440     450     460
55 or34a.pep  AEGKAEDGGS QGADGVRFGF HRVLPFLGVS DGIALRHAVX
   or34-1    AKGKAENG GNQ GADGVRFGF HRVLPFLGVS DGIALRHAVX
      420     430     440     450

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF34 shows 77.6% identity over a 161aa overlap with a predicted ORF (ORF34.ng) from *N. gonorrhoeae*:

orf34.pep

QKLSLRISLWGLGGVFFGVSG LVWFSLGVSXE-----CAC 35

-163-

```

      || |||||:|||||
orf34ng      MMMPFIMLPWIAGVPAVPGQKRLSRISLWGLAGVFFGVSGLVWFSLVGSFSLGVSLGCAC      60
5  orf34.pep      FSGVSFRGSGRGT FVGSTGVSLSVFSACVXGVVRLPVGLSCV-----GRLXXLTRFFLGA      90
      |||||:|||||:|:|:|:|
orf34ng      FSGVSFRGSGWGA FVGSTGVSLSVFSACVP-----VPVNESAARAASEGR--GLTRFFLGA      114
      || |||||:|||||:|||||
orf34.pep      AGDVILLPLSSVPSGCAGSDEAAWWC SGWAASCPTT PFGSQNSVSRGLSVCCGSAXRVLS      150
10 orf34ng      AGDGSPLPLSSVPSGCAGSDEAAWWC SGWAASCPTA PFGSQNSVSRGLSVCCGSVWRVLS      174
      |||||
orf34.pep      S                                                                                      175
orf34ng      PFGLNVLTMPTANAPMAVIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSD      234

```

15 The complete length ORF34ng nucleotide sequence <SEQ ID 213> is:

```

1  ATGATGATGC CGTTCATAAT GCTTCCTTGG ATTGCGGGTG TGCCTGCCGT
51  GCCGGGTCAA AAGAGGTTGT CGAGAATCTC TTTATGGGGT TTGGCCGGCG
101 TGTTCCTCGG GGTGTCCGGT TTGGTATGGT TTTCTTTGGG CGTTTCTTTT
151 TCTTTGGGTG TTTCTTTGGG CTGCGCCTGT TTTTCGGGTG TTTCTTTTCG
201 GGGTTCGGGA TGGGGGCGT TTGTGGGCAG TACGGGGGT TCTTTGAGTG
251 TGTTTTCAGC TTGTGTTCCG GTGCCGGTTA ACGAATCGCG TGCCCGGGCC
301 GCATCCGAAG GCGCGGTTT gACCCGGTTT TTCTTGGGTG CGGCAGGGGA
351 CGGCAGTCCG CTGCGCCTTT CTCTGTGCC GTCCGGCTGT GCGGGTTCGG
401 ATGAGGCGGC GTGTGGTGT TCGGGTTGGG CCGCATCTTG TCCGACGGCG
25 451 CCGTTTGCA GCCAGAATTC GGTTCGCGG GGGCTGTCCG TGTGTTGCGG
501 TTCGGTTTGG AGGTTTGTG CGCCGTTCCG GTTGAATGTG CTGACGATGC
551 CTACTGCCAA TGCCTCGATG GCGGTGATAC AGATGAGCAA TACGGCGCGT
601 ATCAGGAGTT TGGGGGTCAG CCTGAAGGGT TTGTTCGGTT TTTTGCCAT
651 TTTGATTGTG CTTTGGGGT GTCGGGCAAT GCCGTCTGAA GCGGTTTCAG
30 701 ACGGCATTGC CGAGTCAGCG TTGGACGTAG TTTTGGTAGA GGGTAATGAC
751 TTTTGTACG CCGAcgTGG TGCTGACTTT TTGGGTAATC TGCGCTCTGT
801 CTTGCGGGGT GAGGATGCCC ATAACGTAGG TTACATTGCC GTAGGTAATG
851 ATTTTGAGCG GCGCTGTGT AGCGGGCTG ATGCCAGCA GcgtgGCGCG
901 GACTTTGGAC GTGTCCAAG TGTCGCCGCG GATGTCGCCC GCAGTGCGCG
35 951 GCAGGGAGGC GACGGTAATG TAGTTGTATA CGCCTTCGGC GGCCTGTTTC
1001 GAACGTGCAA TCTGACCGAC GAACTGTTT TCGCCTTCGG TGGCGACTTG
1051 TCCGAGCAGC AGCAGGTGGC GGTGTAGCC GACGACGGAG ATTTGGGGCG
1101 TGTAGCCTTT GGTGTGGTTG TTTTGGCGCA GGTAGGAACG GCGGTTGGTT
1151 TCGATACGCA ACGCCATAAC GTgtCATCG GTTtgcgcgc CGGTGGTTcg
40 1201 gCGGTCGATG ACGGATTTTG CGCCGACGCG GGCCCCGCCG ACGACTGCGC
1251 TGAAGCAGCC GCGGAGGCA AGGCTGAGGA CGGCGGCAAT CAGGGTGCGG
1301 ACGGTGTGTG GTTTGGGTTT CATCGGGGAC TTCCTTTCTT GGGCGTTTCA
1351 GACGGCATTG CTTTGCGCCA TGCCGTCTGA

```

This encodes a protein having amino acid sequence <SEQ ID 214>:

```

45 1  MMMPFIMLPW IAGVPAVPGQ KRLSRISLWG LAGVFFGVSG LVWFSLVGSF
51  SLGVSLGCAC FSGVSFRGSG WGA FVGSTGV SLSVFSACVP VPVNESAARA
101 ASEGRGLTRF FLGAAGDGSP LPLSSVPSGC AGSDEAAWWC SGWAASCPTA
151 PFGSQNSVSR GLSVCCGSVW RVLSPFGLNV LTMPTANAPM AVIQMSNTAR
201 IRSLGVSLKG LFGFFAILIV LLGCRAMPSE GS SDGIAESA LDVVLVEGND
50 251 FLYADGGADF LGNLRLFFGG EDAHNVGYIA VGNDFDARLC SGADAQQRGA
301 DFGRVPSVAG DVAR SARQGG DGNVVVYAFG GLFGTCNLTD ELFFAFGGDL
351 SEQQQVAVVA DDGDLGRVAF GLVVL AQVGT GGGFDTORHN VVIGLRAGGS
401 AVDDGFCADG GPADDCAEAA AEGKAEDGGN QGADGVWFGF HRGLPFLGVS
451 DGIALRHAV*

```

55 ORF34ng and ORF34-1 show 90.0% identity in 459 aa overlap:

```

      10      20      30      40      4      50
orf34-1.pep  MMMPFIMLPWIAGVPAVPGONRLSRISLWGLGGVFFGVSGLVWFSLVGS-----LGCAC
      |||||:|||||:|||||:|||||:|||||
60 orf34ng      MMMPFIMLPWIAGVPAVPGQKRLSRISLWGLAGVFFGVSGLVWFSLVGSFSLGVSLGCAC
      10      20      30      40      50      60

      60      70      80      90      100     110
orf34-1.pep  FSGVSFRGSGRGT FVGSTGVSLSVFSACVPASSGCLSVXAVSAGCGLTRFFLGAAGDGSP
      |||||:|||||:|||||:|||||:|||||:|||||
65 orf34ng      FSGVSFRGSGWGA FVGSTGVSLSVFSACVPVPVNESAARAASEGRGLTRFFLGAAGDGSP

```

-164-

		70	80	90	100	110	120
5	orf34-1.pep	120	130	140	150	160	170
	orf34ng	130	140	150	160	170	180
10	orf34-1.pep	180	190	200	210	220	230
	orf34ng	190	200	210	220	230	240
15	orf34-1.pep	240	250	260	270	280	290
	orf34ng	250	260	270	280	290	300
20	orf34-1.pep	300	310	320	330	340	350
	orf34ng	310	320	330	340	350	360
25	orf34-1.pep	360	370	380	390	400	410
	orf34ng	370	380	390	400	410	420
30	orf34-1.pep	420	430	440	450		
	orf34ng	430	440	450	460		

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 26

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 215>:

```

1  ATGAAAACCT TCTTCAAAAC CCTTTCCGCC GCCGCACTCG CGCTCATCCT
51  CGCCGCCTGC GGATT.CAAA AAGACAGCGC GCCCGCCGCA TCCGCTTCTG
101 CCGCCGCCGA CAACGGCGCG GCGTAAAAAA GAAATCGTCT TCGGCACGAC
151 CGTCGGCGAC TTCGGCGATA TGGTCAAAGA ACAATCCAA GCCGAGCTGG
201 AGAAAAAAGG CTACACCGTC AACTGGTTCG AGTTTACCGA CTATGTACGC
251 CCGAATCTGG CATTGGCTGA GGCAGATTG

```

This corresponds to the amino acid sequence <SEQ ID 216; ORF4>:

```

1  MKTFFKTLA AALALILAAC G.QKDSAPAA SASAAADNGA AKKEIVFGTT
51  VGDFGDMVKE IQAELEKKG YTVKLVEFTD YVRPNLALAE GEL

```

Further sequence analysis revealed the complete nucleotide sequence <SEQ ID 217>:

```

1  ATGAAAACCT TCTTCAAAAC CCTTTCCGCC GCCGCACTCG CGCTCATCCT
55  51  CGCCGCCTGC GCGGTCAAAA AAGACAGCGC GCCCGCCGCA TCCGCTTCTG
101 CCGCCGCCGA CAACGGCGCG GCGAAAAAAG AAATCGTCTT CGGCACGACC
151 GTCGGCGACT TCGGCGATAT GGTCAAAGAA CAAATCCAAG CCGAGCTGGA
201 GAAAAAAGGC TACACCGTCA AACTGGTTCGA GTTTACCGAC TATGTACGCC

```

20

15

20

**Homology with a predicted ORF from *N.meningitidis* (strain A)**

25

30

35

40

45

50

55

This is predicted to encode a protein having amino acid sequence <SEO ID 220>:

1 MKTFFKTLSA AALALILAAC GGQKDSAPAA SASAAADNGA AXKEIVEGTT

-166-

51 VGDFGDMVKE XIQPELEKKG YTVKLVESTD YVRXNLALAE GELDINXVQH  
 101 XXYLDDKXKX HNLDITXVXQ VPTAPLGLYP GKLSLXXVK XGSTVSAPND  
 151 PXXFXRVLMV LDELGXIKLK DXIXXXXXXX XXXXXXXXXX XXXXXXXXXX  
 201 XXXXAXXXXX XXXXXXXXXX GMKLTEALFQ EPSFAYVNWS AVKTADKDSQ  
 251 WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAK\*

A leader peptide is underlined.

Further analysis of these strain A sequences revealed the complete DNA sequence <SEQ ID 221>:

1 ATGAAACCT TCTTCAAAAC CCTTTCGCC GCCGCACTCG CGCTCATCCT  
 51 CGCGCCTGC GGCGGTCAAA AAGATAGCGC GCCCGCCGCA TCCGCTTCTG  
 101 CCGCGCCGA CAACGGCGCG GCGAAAAAG AAATCGTCTT CGGCACGACC  
 151 GTCGGCGACT TCGCGGATAT GGTCAAAGAA CAAATCCAAC CGGAGCTGGA  
 201 GAAAAAGGC TACACCGTCA AACTGGTCGA GTTACCGAC TATGTGCGCC  
 251 CGAATCTGGC ATTGGCTGAG GCGGAGTTGG ACATCAACGT CTTCCAACAC  
 301 AAACCTATC TTGACGACTT CAAAAAGAA CACAATCTGG ACATACCGA  
 351 AGTCTTCCAA GTGCCGACCG CGCCTTTGGG ACTGTACCGG GGCAAGCTGA  
 401 AATCGCTGGA AGAAGTCAAA GACGGCAGCA CCGTATCCGC GCCCAACGAC  
 451 CCGTCCAAC TCGCCCGCGT CTTGGTGATG CTCGACGAAC TGGGTGGAT  
 501 CAAACTCAA GACGGCATCA ATCCGCTGAC CGCATCCAAA GCGGACATTG  
 551 CCGAAACCT GAAAAACATC AAAATCGTCG AGCTTGAAGC CGCGCAACTG  
 601 CCGCGTAGCC GCGCCGAGCT GGATTTTGCC GTCGTCAACG GCAACTACGC  
 651 CATAAGCAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT  
 701 TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA  
 751 TGGCTTAAAG ACCTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA  
 801 CGCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGAATGAAG  
 851 GCGCAGCCAA ATAA

This encodes a protein having amino acid sequence <SEQ ID 222; ORF4a-1>:

1 MKTFFKTLSA AALALILAAC GGQKDSAPAA SASAAADNGA AKKEIVFGTT  
 51 VGDFGDMVKE QIQPELEKKG YTVKLVEFTD YVRPNLALAE GELDINVFQH  
 101 KPYLDDFKKE HNLDITEVFQ VPTAPLGLYP GKLSLEEVK DGSTVSAPND  
 151 PSNFARVLVM LDELGWIKLK DGINPLTASK ADIAENLNKI KIVELEAAQL  
 201 PRSRADVDFV VVNGNYAISS GMKLTEALFQ EPSFAYVNWS AVKTADKDSQ  
 251 WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAK\*

ORF4a-1 and ORF4-1 show 99.7% identity in 287 aa overlap:

35	orf4a-1	10	20	30	40	50	60
		MKTFFKTLSA <u>AALALILAAC</u> GGQKDSAPAAASASAAADNGAAKKEIVFGTTVGDFGDMVKE					
	orf4-1	MKTFFKTLSA <u>AALALILAAC</u> GGQKDSAPAAASASAAADNGAAKKEIVFGTTVGDFGDMVKE					
40	orf4a-1	70	80	90	100	110	120
		QIQPELEKKG YTVKLVEFTD YVRPNLALAE GELDINVFQH KPYLDDFKKE HNLDITEVFQ					
	orf4-1	QIQAELEKKG YTVKLVEFTD YVRPNLALAE GELDINVFQH KPYLDDFKKE HNLDITEVFQ					
45	orf4a-1	130	140	150	160	170	180
		VPTAPLGLYP GKLSLEEVK DGSTVSAPND PSNFARVLV MLDELGWIKLK DGINPLTASK					
	orf4-1	VPTAPLGLYP GKLSLEEVK DGSTVSAPND PSNFARVLV MLDELGWIKLK DGINPLTASK					
50	orf4a-1	190	200	210	220	230	240
		ADIAENLNKI KIVELEAAQL PRSRADVDF VVNGNYAISS GMKLTEALFQ EPSFAYVNWS					
	orf4-1	ADIAENLNKI KIVELEAAQL PRSRADVDF VVNGNYAISS GMKLTEALFQ EPSFAYVNWS					
55	orf4a-1	250	260	270	280		
		AVKTADKDSQ WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAKX					
	orf4-1	AVKTADKDSQ WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAKX					
60		250	260	270	280		

```

5      lip2.pasha          10      20
                        MNFKLLGVALVSALALTACKDEKAQAP-----
                        |||::|| |||:|| :| |
ORF4      VXTPNPDGRTPCPSFLFETATTSGENMKTFKTL$AAAL--ALILAACGFKKTARPPHPL
                110        120        130        140        150

10     lip2.pasha          30      40      50      60      70      80
                    -ATTAKTENKAPLKVGVMTGPEAQMTEVAVKIAKEKYGLDVELVQFTEYQTQPNAALHSKD
                      ::| : |: :| :::: :| || | :||:|:|:|:|:| :| :
ORF4      LPPPTTARRRKEIVFGTTVGGDFGDMVKEIQAELEKKGYTVKLVEFTDYVRPNLALAEGE
                160        170        180        190        200        210

15     lip2.pasha          90      100      110      120      130      140
                    LDANAFQTVPYLEQEVKDRGYKLAIIGNTLVWPPIAAYSKKIKNISELKGATVAIPNNAS
                    |
ORF4      L.....

```

ORF4 shows 93.6% identity over a 94aa overlap with a predicted ORF (ORF4.ng) from *N. gonorrhoeae*:

			10	20	30
25	orf4nm.pep	MKTFFKTL	SAAALALILAACGXQKDSAPAA		
	orf4ng	RANAVXTPNPDGRTPCLSLFETATTSGENMKTFFKTLSTASLALILAACGGGQKDSAPAA	:   :		
		200	210	220	230
				240	250
30	orf4nm.pep	SASA-AADNGA	AKKEIVFGTTVGDFGDMVKEIQAELEKKGYTVKLVEFTDYVRPNLALA		
	orf4ng	SAAAPSADNGA	AKKEIVFGTTVGDFGDMVKEIQAELEKKGYTVKLVEFTDYVRPNLALA		
		260	270	280	290
				300	310
35	orf4nm.pep	EGEL			
	orf4ng	EGELDINVFQH	KPYLD DFKKEHNL DITEAFQVP TAPLG LYPGLKSLEE VKDGSTVS APN		
		320	330	340	350
				360	370

40 The complete length ORF4ng nucleotide sequence <SEQ ID 223> was predicted to encode a protein having amino acid sequence <SEQ ID 224>:

	1	MKTFEKTLLST	ASLALILAAC	GGQKDSAPAA	SAAAPSDADNG	AAKKEIVFGT
	51	TVGDFGDMVK	EIQIQAELKK	GYTVKLVEFT	VDYVRNLALA	EGDGLDNVFG
45	101	HKPYLDDFEAK	EHNLDITEAF	QVPTAPLGLY	PGKLKSLLEAV	KGDGSTVSAPN
	151	DPSNFARALV	MLNELGWIKL	KDGINPLTAS	KADIAENLKN	IKIVELEAAQ
	201	LPRSRADVDF	AVVNGNYAIS	SGMKLTEALF	QEPSFAYVNW	SAVKTADKDS
	251	QWLKDVTVEY	NSDAFKAYAH	KRFEGYKYPA	AWNEGAAK*	

Further analysis revealed the complete length ORF4ng DNA sequence <SEQ ID 225> to be:

50	1	atgAAAACCT	TCTTCAAAAC	cctttccgcc	gccgcaCTCG	CGCTCATCCT
	51	CGCAGCCTGc	ggCggtcaAA	AAGACAGCGC	GCCCgcagcc	tctgcCGCCG
	101	CCCCTTCTGC	CGATAACGgc	gCgGCGAAAA	AAGAAAtcgt	ctTCGGCAGC
	151	Accgtgggcg	acttcggcgA	TatggtCAAA	GAACAAATCC	AagcCGAgct
	201	gGAGAAAAA	GgctACACcg	tcAAattggt	cgaatttacc	gactatgtGC
	251	gCCGGAATCT	GGCATTTGGCG	GAGGGCGAGT	TGACATCAA	CGTCTTCCAA
55	301	CACAAACCCT	ATCTTGACGA	TTTCAAAAAA	GAACACAACC	TGGACATCAC
	351	CGAAGCCTTC	CAAGTGCCGA	CCGCGCCTTT	GGGACTGTAT	CCGGGCAAAC
	401	TGAAATCGCT	GGAAGAAGTC	AAAGACGGCA	GCACCGTATC	CGCGCCCAac
	451	gACccgTCCA	ACTTCGCACG	CGCCTTGGTG	ATGCTGAAC	AACTGGGTTG
	501	GATCAAATCT	AAAGACGGCA	TCAATCCGCT	GACCGCATCC	AAAGCCGACA
60	551	TCGCGGAAAA	CCTGAAAAAC	ATCAAAATCG	TCGAGCTTGA	AGCCGCAAAA

This encodes a protein having amino acid sequence <SEQ ID 226; ORF4ng-1>:

10

**This shows 97.6% identity in 288 aa overlap with ORF4-1:**

15

20

25

30

35

40

45

50

60



-169-

	lip2_pasha	TEVAVKIAKEKYGLDVELVQFTEYTPQNAALHSDKLDANAFQTPYLEQEVKDRGYKLAI	60	70	80	90	100	110
5	orf4ng-1.pep	AFQVPTAPLGLYPGKLKSLSEEVKDGSTVSAPNDPSNFARALVMLNELGWIKLKDGINPLT	120	130	140	150	160	170
	lip2_pasha	IGNTLVWPIAAYSKKIKNISELKDGATVAIPNNASNTARALLLLOAHGLLKLKDPKN-VF	120	130	140	150	160	170
10	orf4ng-1.pep	ASKADIAENLKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTE--ALFQEPSFA	180	190	200	210	220	230
	lip2_pasha	ATENDI IENPKNIKIVQADTSLLRMLDDVELAVINNTYAGQAGLSPDKGIIVESKDSP	180	190	200	210	220	230
15	orf4ng-1.pep	YVNWSAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKYPAAWNEGAAXX	240	250	260	270	280	289
	lip2_pasha	YVNLVVSREDNKDDPRLQTFVKSFQTEEVFQEALKLFNGGVVKGW	240	250	260	270		

Based on this analysis, including the homology with the outer membrane protein of *Pasteurella haemolítica*, and on the presence of a putative prokaryotic membrane lipoprotein lipid attachment site in the gonococcal protein, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF4-1 (30kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figures 8A and 8B show, respectively, the results of affinity purification of the His-fusion and GST-fusion proteins. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result), Western blot (Figure 8C), FACS analysis (Figure 8D), and a bactericidal assay (Figure 8E). These experiments confirm that ORF4-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 8F shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF4-1.

### Example 27

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 227>:

	1	CCTCGTCGTC	CTCGGCATGC	TCCAGTTTCA	AGGGGCGATT	TACTCCAAGG
	51	CGGTGGAACG	TATGCTCGGC	ACGGTCATCG	GGCTGGGCGC	GGGTTTGGGC
40	101	GTTTTATGGC	TGAACAGCA	TTATTTCAC	GGCAACCTCC	TCTTCTACCT
	151	CACCGTCGGC	ACGGCAAGCG	CACCTGGCCG	CTGGGCGGCG	GTGGGCAAAA
	201	ACGGCTACGT	CCCTmTGCTG	GCAGGGCTGA	CGATGTGTAT	GCTCATCGGC
	251	GACAACGGCA	GCGAATGGCT	CGACAGCGGA	CTCATGCGCG	CCATGAACGT
	301	CCTCATCGGC	GyGGCCATCG	CCATCGCCGC	CGCCAACTG	CTGCCGCTGA
	351	AATCCACACT	GATGTGGCGT	TTCATGCTTG	CCGACAACCT	GGCCGACTGC
45	401	AGCAAAATGA	TTGCCGAAAT	CAGCAACGGC	AGGCGCATGA	CCCGCGAACG
	451	CCTCGAGGAG	AACATGGCGA	AAATGCGCCA	AATCAACGCA	CGCATGGTCA
	501	AAAGCCGCAG	CCATCTCGCC	GCCACATCGG	GCGAAAGCTG	CATCAGCCCC
	551	GCCATGATGG	AAGCCATGCA	GCACGCCAC	CGTAAATCG	TCAACACCAC
	601	CGAGCTGCTC	CTGACCACCG	CCGCCAAGCT	GCAATCTCCC	AACTCAACG

5 651 GCAGCGAAAT CCGGCTGCTT GACCGCCACT TCACACTGCT CCAAAC....  
 701 ..... GC AGACACGCCC GCCGCATCCG  
 751 CATCGACACC GCCATCAACC CCGAACTGGA AGCCCTCGCC GAACACCTCC  
 801 ACTACCAATG GCAGGGCTTC CTCTGGCTCA GCACCGATAT GCGTCAGGAA  
 851 ATTTCCGCCC TCGTCATCCT GCTGCAACGC ACCCGCCGCA AATGGCTGGA  
 901 TGCCACGAA CGCCAACACC TGCGCCAAG CCTGCTTGA

This corresponds to the amino acid sequence <SEQ ID 228; ORF8>:

10 1 .....PRRP RHAPVSRGDL LQGGTYARH GHRAGRGFGR FMAEPALFPR  
 51 QPPLLPHRRH GKRTGRLGGG RQKRLRPXAG RADDVYAHRR QRQRMARQT  
 101 HARHERPHRR GHRHRRRQTA AAEIHTDVAF HACRQPGRLQ QNDCRNQQRQ  
 151 AHDPRTPRGE HGENAPNQT HGQKPQPSRR HIGRKLHQPR HDGSHAARPP  
 201 XNRQHHRAAP DHRRAAISQ TQRQNPAAQ PPLHTAPN.. .....Q  
 251 TRPPHPRHRH HQPRTGSPRR TPPLPMAGLP LAQHRYASGN FRPRHPAATH  
 301 PPQMAGCPRT PTPAPKPA\*

15 Computer analysis of this amino acid sequence gave the following results:

#### Sequence motifs

ORF8 is proline-rich and has a distribution of proline residues consistent with a surface localization. Furthermore the presence of an RGD motif may indicate a possible role in bacterial adhesion events.

#### 20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF8 shows 86.5% identity over a 312aa overlap with a predicted ORF (ORF8.ng) from *N. gonorrhoeae*:

25 orf8ng 1 MDRDDLRLRRPHAPVPRRDLLQGGTYARYGHRAGRGFGRFMAEPALFPR 50  
 orf8.pep 1 .....PRRERHAPVSRGDLQGGTYARHGHRAGRGFGRFMAEPALFPR 44  
 30 orf8ng 51 QPLLPDHRHGKRTGRLGGGRQKRLRPYVGADDVHAHRRQRQRMARQRP 100  
 orf8.pep 45 QPLLPHRRHGKRTGRLGGGRQKRLRPXAGRADDVYAHRRQRQRMARQRT 94  
 orf8ng 101 DARDERPHRRRHRHCRQTAAAEIHTDVAFHACRQPGRLQONDCRNQQRQ 150  
 orf8.pep 95 HARHERPHRRRGHRHRRRQTA AAEIHTDVAFHACRQPGRMQONDCRNQQRQ 144  
 35 orf8ng 151 AYDARTFGAEYQONAPNQRTHGQKPQPPRRHIGRKPQPLHDGSHAARPP 200  
 orf8.pep 145 AHDPRTPRGEHGENAPNQRTHGQKPQPSRRHIGRKLHQPRHDGSHAARPP 194  
 40 orf8ng 201 QNRQHHRAAPDHRRAAISQTQRQNPAAQ PPLHTAPNRPATNRRPHQRQ 250  
 orf8.pep 195 XNRQHHRAAPDHRRAAISQTQRQNPAAQ PPLHTAPN.....Q 244  
 45 orf8ng 251 TRPPHPRHRHQPRTGSPRRTPPLPMAGFPLAQHRYASGNFRPRHPATH 300  
 orf8.pep 245 TRPPHPRHRHQPRTGSPRRTPPLPMAGLPLAQHRYASGNFRPRHPAATH 294  
 orf8ng 301 PPQMAGCPRTPTPAPKPA\* 319  
 orf8.pep 295 PPQMAGCPRTPTPAPKPA\* 313

50 The complete length ORF8ng nucleotide sequence <SEQ ID 229> is predicted to encode a protein having amino acid sequence <SEQ ID 230>:

55 1 MDRDDLRLRR RHAPVPRDL LQGGTYARY GHRAGRGFGR FMAEPALFPR  
 51 QPLLPDHRH GKRTGRLGGG RQKRLRPYVG GADDVHAHRR QRQRMARQRP  
 101 DARDERPHRR RHRHCRQTA AAEIHTDVAF HACRQPGRLQ QNDCRNQQRQ  
 151 AYDARTFGAE YQONAPNQT HGQKPQPPRR HIGRKPQPL HDGSHAARPP

201 QNRQHHRAAP DHRROAAISQ TQRQNPAAAR PPLHTAPNRP ATNRRPHQRQ  
 251 TRPPHPRHR HQPRTGSPRR TPPLPMAGFP LAQHQYASGN FRPRHPPATH  
 301 PPQMACFPRT PTPAPKPA\*

Based on the sequence motifs in these proteins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 28

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 231>:

```

10      1  ..GAAATCAGCC TCGGTTCCGA CNACAGGCCG GTTTCCTGN CGAAGCGGCG
      51  GGATTCGGAA CGTTTCTGCG TGTTGGACGG CGGCAACAGC CGGCTCAAGT
     101  GGGCGTGGGT GGAAAACGGC ACGTTCGCAA CCGTCGGTAG CGCGCCGTAC
     151  CGCGATTGTG CGCCTTTGGG CGCGGAGTGG GCGGAAAAGG CGGATGGAAA
     201  TGTCCGCATC GTCGGTTGCG CTGTGTGCGG AGAATTCAAA AAGGCACAAG
     251  TGCAGGAACA GTCGCCCCGA AAAATCGAGT GGCTGCCGTC TTCCGCACAG
     301  GCTTT.GGCA TACGCAACCA CTACCGCCAC CCCGAAGAAC ACGGTTCCGA
     351  CCGCTGGTTC AACGCCTTGG GCAGCCGCGG CTTCAGCCGC AACGCCTGCG
     401  TCGTCGTCAG TTGCGGCACG GCGGTAACGG TTGACGCGCT CACCGATGAC
     451  GGACATTATC TCGGAGA.GG AACCATCATG CCCGGTTTCC ACCTGATGAA
     501  AGAATCGCTC GCCGTCCGAA CCGCCAACCT CAACCGGCAC GCCGGTAAGC
     551  GTTATCCTTT CCCGACCGG..
  
```

This corresponds to the amino acid sequence <SEQ ID 232; ORF61>:

```

25      1  ..EISLRSDXRP VSVXKRRDSE RFLLLDGGNS RLKAWVENG TFATVGSAPY
      51  RDLSPGAEW AEKADGNVRI VGCVCGEFK KAQVQEQLAR KIEWLPSSAQ
     101  AXGIRNHYRH PEEHSGDRWF NALGSRRESR NACVVVSCGT AVTVDALTD
     151  GHYLGXGTIM PGFHLMKESL AVRTANLNRH AGKRYPFPT..
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 233>:

```

30      1  ATGACGGTTT TGAAGCTTTC GCACTGGCGG GTGTTGGCGG AGCTTGCCGA
      51  CGGTTTGCCG CAACACGTCT CGCAACTGGC GCGTATGGCG GATATGAAGC
     101  CGCAGCAGCT CAACGGTTTT TGGCAGCAGA TGCCGGCGCA CATACGCGGG
     151  CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CATTGGCGGT
     201  TTTCGATGCC GAAGGTTTGC GCGAGCTGGG GGAAAGGTCT GGTTTTCAGA
     251  CGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
     301  GCGCGGATTG CGCCGGACAA GGCGCACAAA ACCATATGCG TGACCCACCT
     351  GCAAAGTAAG GGCAGGGGGC GGCAGGGGCG GAAGTGGTCG CACCGTTTGG
     401  GCGAGTGTCT GATGTTCACT TTTGGCTGGG TGTTTGACCG GCCGAGTAT
     451  GAGTTGGGTT CGTGTCCGCC GTTTCGGGCA GTGGCGTGTC GGCGCGCCTT
     501  GTCGCGTTTA GGTTTGGATG TGCAGATTAA GTGGCCCAAT GATTGGTTTG
     551  TCGGACGCGA CAAATTGGGC GGCATTCTGA TTGAAACGGT CAGGACGGGC
     601  GGCAAAACGG TTGCCGTGGT CCGTATCGGC ATCAATTTTG TCCTGCCCAA
     651  GGAAGTAGAA AATGCCGCTT CCGTGCAATC GCTGTTTCTG ACGGCATCGC
     701  GGCGGGGCAA TGCCGATGCC GCCGTGCTGC TGGAAACGCT GTTGGTGGAA
     751  CTGGACGCGG TGTTGTTGCA ATATGCGCGG GACGGATTTC CGCCTTTTGT
     801  GGCGGAATAT CAGGCTGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT
     851  TGCGCGACGG CGAAACCGTG TTCGAAGGCA CGGTTAAAGG CGTGACGGA
     901  CAAGGCGTTT TGCATTGGA AACGGCAGAG GGCAAAACAG CCGTCGTCAG
     951  CGGCGAAATC AGCCTGCGGT CCGACGACAG GCCGGTTTCC GTGCCGAAGC
    1001  GGCGGGATTG GGAACGTTTT CTGCTGTTGG ACGGCGGCAA CAGCCGGCTC
    1051  AAGTGGGCGT GGGTGGAAAA CGGCACGTTT GCAACCGTCG GTAGCGCGCC
    1101  GTACCGCGAT TTGTCGCCTT TGGGCGCGGA GTGGGCGGAA AAGGCGGATG
    1151  GAAATGTCCG CATCGTCGGT TGCCTGTGTG GCGGAGAATT CAAAAGGCA
    1201  CAAGTGCAGG AACAGCTCGC CCGAAAAATC GAGTGGCTGC CGTCTTCCGC
    1251  ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCCGAA GAACACGGTT
    1301  CCGACCGCTG GTTCAACGCC TTGGGCAGCC GCCGCTTCTG CCGCAACGCC
    1351  TCGCTCGTCG TCACTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCGA
    1401  TGACGGACAT TATCTCGGGG GAACCATCAT GCCCGGTTTC CACCTGATGA
    1451  AAGAATCGCT CGCCGTCCGA ACCGCCAACC TCAACCGGCA CGCCGGTAAG
    1501  CGTTATCCTT TCCCAGCAC AACGGGCAAT GCCGTCGCCA GCGGCATGAT
    1551  GGATGCGGTT TCGGCTCGG TTATGATGAT GCACGGGCGT TTGAAAGAAA
    1601  AAACCGGGGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CGGCGGCGCG
  
```

1651 GCAAAAGTTG CCGAAGCCCT GCCGCGTGCA TTTTGGCGG AAAATACCGT  
1701 GCGCGTGGCG GACAACTCG TCATTTACGG GTTGTTGAAC ATGATTGCCG  
1751 CCGAAGGCAG GGAATATGAA CATATTTAA

**This corresponds to the amino acid sequence <SEQ ID 234; ORF61-1>:**

5	1	MTVLKLSHRW	VLAELADGLP	QHVSQALRMA	DMKPOQLNGF	WQOMPAPIRG
	51	LLRQHDGYWR	LVRPLAVFDA	EGRLRELGERS	GFQTALKHEC	ASSNDEILEL
	101	ARIAPDKAHK	TICVTHLQSK	GRGRQGRKWS	HRLGECLMFS	FGWVFDPRQY
	151	ELGSLSPVAA	VACRRALSR	GLDVQIKWPN	DLVVGRDKLG	GILIVTRTGT
	201	GKTVAUVGIG	INFVLPKEVE	NAASVQSLFQ	TASRRGNADA	AVLLETLLE
10	251	LDVALLQYAR	DGFAPFVAEY	QAANRDHGKA	VLLLRDGETV	FEGTVKGVGD
	301	QGVLEHLETA	GKQTVVSGEI	SLRSDDRPVS	VPKRRDSERF	LLLDGGSNSRL
	351	KWAWVENGTF	ATVGSAPYRD	LSPLGAEWAE	KADGNVRIVG	CAVCGEFKKA
	401	QVQEQLARKI	EWLPSAQAAL	GIRNHYRHPE	EHGSDRWfNA	LGSRRFSRfNA
	451	CVVVSCTAV	TVDALTPDDG	YLGGTIMPGF	HLMKESLAVR	TANLNHRHAGK
15	501	RYFFPTTTGN	AVASGMMDAV	CGSVMMMHGR	LKEKTGAGKP	VDVLIITGGGA
	551	KYFAEALPPA	FLAENTVRVA	DNLVIYGLLN	MIAAEGREYE	HI*

Figure 9 shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF61-1. Further computer analysis of this amino acid sequence gave the following results:

Homology with the baf protein of *B. pertussis* (accession number U12020).

20 ORF61 and baf protein show 33% aa identity in 166aa overlap:

	orf61	23	LLLDGGSRLKWAWVE-NGTFATVGSAPYR----DLSPGLAEWAEKADGNVRIVGCACVG	77
			+L+D GNSRLK W + + A AP DL LG A R +G V G	
	baf	3	ILIDSGNSRLKVGWFDPDAPQAAREPAPVAFDNLDDLALGRWLATLPRRPQRALGVNVAG	62
25	orf61	78	EFKKAQVQEQLAR---KIEWLPSSAQAXGIRNHYRHPEEHGSDRW---FNALGSRRFRSN	131
			+ + L I WL + A G+RN YR+P++ G+DRW L +	
	baf	63	LARGEIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDLQGLADRWACMVGVLARQPSVHP	122
30	orf61	132	ACVVVSCGTAVTVDALTDGHYLGXGTIMPGFHLMKESLAVRTANL	177
			+V S GTA T+D + D + G G I+PG +M+ +LA TA+L	
	baf	123	PLLVASFGTATTLDITIGPDNVFPG-GLILPGPAMMRGALAYGTAHL	167

**Homology with a predicted ORF from *N.meningitidis* (strain A)**

ORF61 shows 97.4% identity over a 189aa overlap with an ORF (ORF61a) from strain A of *N.*

35 *meningitidis:*

	orf6l.pep	EISLRSDXRPPVSXKRRDSEFLLLDGGS 	10                  20                  30
40	orf6la	TVFEGTVKGVDGQGVHLLETAEGKQT VVSGEISLRSDDRPVSVPKR RDSERFLLLDGGS 290              300              310              320              330              340	
	orf6l.pep	RLKWAVVENGTTFATVGSA PYRDLSP LGAWEAKADGNVRIVGC AVCGEFKKAQVQEQLAR 	40            50            60            70            80            90
45	orf6la	RLKWAVVENGTTFATVGSA PYRDLSP LGAWEAEKV DGNVRIVGC AVCGEFKKAQVQEQLAR 350              360              370              380              390              400	
50	orf6l.pep	KIEWLPSSAQAXGIRNH YRHPEEHGSDRW FNFALGSRRF SRNACVVVSC GTAVTVDALTDD 	100            110            120            130            140            150
	orf6la	KIEWLPSSA QALGIRNH YRHPEEHGSDRW FNFALGSR RFNRNACVVV SCGTAVTVDALTDD 410            420            430            440            450            460	
55	orf6l.pep	GHYLG XGTIMP GFHLMKE SLAVRTAN LN RHAG KRYPFPT 	160            170            180            189
	orf6la	GHYL G-GTIM PGFHLM KE SLAVRTAN LN RHAG KRYPFPTTTGN AVASGMMD A VC GSVM MM 470            480            490            500            510            520	
60	orf6la	HGR LKEKTG AGKPVDVIITGG GA AKVA EAL PPA FLA ENT VRV ADNL VIHG LLNL IA EG GG	

530 540 550 560 570 580

The complete length ORF61a nucleotide sequence <SEQ ID 235> is:

```

1  ATGACGGTTT TGAAGCCTTC GCACTGGCGG GTGTTGGCGG AGCTTGCCGA
5  51  CGGTTTGCCG CAACACGTCT CGCAACTGGC GCGTATGGCG GATATGAAGC
    101  CGCAGCAGCT CAACGGTTTT TGGCAGCAGA TGCCGGCGCA CACACGCGGG
    151  CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CATTGGCGGT
    201  TTTGATGCC GAAGGTTTGC GCGAGCTGGG GGAAAGGTCG GGTTTTCAGA
    251  CGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
    301  GCGCGGATTG CGCCGGACAA GGCGCACAAA ACCATATGTG TGACCCACCT
10  351  GCAAAGTAAG GGCAGGGGGC GGCAGGGGCG GAAGTGGTCG CACCGTTTGG
    401  GCGAGTGTCT GATGTTCACT TTTGGCTGGG TGTTTGACCG GCCGCAGTAT
    451  GAGTTGGGTT CGCTGTCGCC TGTTGCGGCA GTGGCGTGCC GGCGCGCCTT
    501  GTCGCGTTTG GGTGTTGAAA CGCAAATCAA GTGGCCAAAC GATTTGGTCTG
    551  TCGGACGCGA CAAATTTGGC GGCATTCTGA TTGAAACGGT CAGGACGGGC
15  601  GGCAAACGG TTGCCGTGGT CCGTATCGGC ATCAATTTCC TGCTGCCCAA
    651  GGAAGTGGAA AACGCCGCTT CCGTGCAATC GCTGTTTCAG ACGGCATCGC
    701  GCGGGGAAA TGCCGATGCC GCCGTGTTGC TGGAAACGCT GTTGGCGGAA
    751  CTTGATGCGG TGTTGTTGCA ATATGCGCGG GACGGATTTG CGCCTTTTGT
    801  GGCGGAATAT CAGGCTGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT
20  851  TGCGCGACGG CGAAACCGTG TTCGAAGGCA CGGTTAAAGG CGTGGACGGA
    901  CAAGGCGTTC TGCACTTGA AAGCGCAGAG GGCAAACAGA CCGTTCGTAG
    951  CCGCGAAATC AGCCTGCGGT CCGACGACAG GCCGGTTTCC GTGCCGAAGC
1001  GGCGGGATTC GGAACGTTTT CTGCTGTTGG ACGCGGGCAA CAGCCGGCTC
    1051  AAGTGGGCGT GGTGGAATAA CGGCACGTTT GCAACCGTCG GTAGCGCGCC
25  1101  GTACCGCGAT TTGTCGCCCT TGGGCGCGGA GTGGGCGGAA AAGGTGGATG
    1151  GAAATGTCCG CATCGTCGCT TCGCGCGTGT GCGGAGAATT CAAAAGGCA
    1201  CAAGTGCAGG AACAGCTCGC CCGAAAAATC GAGTGGCTGC CGTCTTCCGC
    1251  ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCGAA GAACACGGTT
    1301  CCGACCGCTG GTTCAACGCC TTGGGCGAGC GCCGCTTCAG CCGCAACGCC
30  1351  TGCGTCGTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCAG
    1401  TGACGGACAT TATCTCGGGG GAACCATCAT GCCCGGTTT CACCTGATGA
    1451  AAGAATCGCT CGCCGTCCGA ACCGCCAACC TCAACCGGCA CGCCGGTAAG
    1501  CGTTATCCTT TCCCGACCAC AACGGGCAAT GCCGTCGCCA GCGGCATGAT
    1551  GGATGCGGTT TGCGGCTCGG TTATGATGAT GCACGGGCGT TTGAAAGAAA
35  1601  AAACCGGGGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CGGCGGCGCG
    1651  GCAAAGTTG CCGAAGCCCT GCCGCCTGCA TTTTGGCGG AAAATACCGT
    1701  GCGCGTGGCG GACAACCTCG TCATTACCGG GCTGCTGAAC CTGATTGCCG
    1751  CCGAAGGCGG GGAATCGGAA CATACTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 236>:

```

40  1  MTVLKPSHWR VLAELADGLP OHVSQLARMA DMKPQQLNGF WQOMPAHIRG
    51  LLRQHDGYWR LVRPLAVFDA EGLRELTERS GFQTALKHEC ASSNDEILEL
    101  ARIAPDKAHK TICVTHLQSK GRGRQGRKWS HRLGECMFES FGWVFDPRQY
    151  ELGSLSPVAA VACRRALSRL GLKTQIKWPN DLVVGRDKLG GILIIETVRTG
45  201  GKTAVVVGIG INFVLPKEVE NAASVQSLFQ TASRRGNADA AVLLETLLAE
    251  LDAVLLQYAR DGFAPFVAEY QAANRDHGKA VLLLRDGETV FEGTVKGVDG
    301  QGVLEHLETA GKQTVVSGEI SLRSDRPPVS VPKRRDSERF LLLDGGNSRL
    351  KWAWVENGTF ATVGSAPYRD LSPLGAEWAE KVDGNVRIVG CAVCGEFKKA
    401  QVQEQIARKI EWLPSAQAL GIRNHYRHPE EHGSDFWFNA LGSRRFSRNA
    451  CVVVSCGTAV TVDALTDGHH YLGGTIPMGF HLMKESLAVR TANLNRHAGK
50  501  RYFPFTTTGN AVASGMMDAV CGSVMMMHGR LKEKTGAGKP VDVIITGGGA
    551  AKVAEALPPA FLAENTVRVA DNLVIHGLLN LIAAEGGESE HT*

```

ORF61a and ORF61-1 show 98.5% identity in 591 aa overlap:

```

55  orf61a.pep  10 20 30 40 50 60
    MTVLKPSHWRVLAELADGLPQHVSQALARMADMKPQQLNGFWQOMPAHIRGLLRQHDGYWR
    orf61-1     10 20 30 40 50 60
    MTVLKLSHWRVLAELADGLPQHVSQALARMADMKPQQLNGFWQOMPAHIRGLLRQHDGYWR

60  orf61a.pep  70 80 90 100 110 120
    LVRPLAVFDAEGLRELTERSFGQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK
    orf61-1     70 80 90 100 110 120
    LVRPLAVFDAEGLRELTERSFGQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK

65  130 140 150 160 170 180

```

	orf61a.pep	GRGRQGRKWSHRLGECLMFSFGWVDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN	
	orf61-1	GRGRQGRKWSHRLGECLMFSFGWVDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN	
5		130 140 150 160 170 180	
	orf61a.pep	190 200 210 220 230 240	
	orf61-1	DLVVGRDKLGGILIIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA	
10		190 200 210 220 230 240	
	orf61a.pep	250 260 270 280 290 300	
	orf61-1	AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG	
15		250 260 270 280 290 300	
	orf61a.pep	310 320 330 340 350 360	
	orf61-1	QGVHLHLETAEGKQTVVSGEISLRSDRPPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF	
20		310 320 330 340 350 360	
	orf61a.pep	370 380 390 400 410 420	
	orf61-1	ATVGSAPYRDLSPGLAEWAEEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL	
25		370 380 390 400 410 420	
	orf61a.pep	430 440 450 460 470 480	
	orf61-1	GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDGHHYLGGTIMPGF	
30		430 440 450 460 470 480	
	orf61a.pep	490 500 510 520 530 540	
	orf61-1	HLMKESLAVRTANLNRHAGKRYFPPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP	
35		490 500 510 520 530 540	
	orf61a.pep	550 560 570 580 590	
	orf61-1	VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNIAAEGGESEHTX	
40		550 560 570 580 590	
	orf61a.pep	550 560 570 580 590	
	orf61-1	VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNIMIAAEGREYEHIX	
45		550 560 570 580 590	

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF61 shows 94.2% identity over a 189aa overlap with a predicted ORF (ORF61.ng) from *N.*

50 *gonorrhoeae*:

	orf61.pep	EISLRSDXRPVSVXKRRDSERFLLLDGGNS	30
	orf61ng	TVCEGTVKGVDGRGVLHLETAEGEQTVVSGEISLRPDNRSVSVKRPDSERFLLLEGNS	211
55	orf61.pep	RLKWAWVENGTFATVGSAPYRDLSPGLAEWAEEKADGNVRIVGCAVCGEFKKAQVQEQLAR	90
	orf61ng	RLKWAWVENGTFATVGSAPYRDLSPGLAEWAEEKADGNVRIVGCAVCGESKKAQVKEQLAR	271
60	orf61.pep	KIEWLPSSAQAXGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTD	150
	orf61ng	KIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTD	331
	orf61.pep	GHYLGXGTIMPGFHLMKESLAVRTANLNRHAGKRYFPPT	189
65	orf61ng	GHYLG-GTIMPGFHLMKESLAVRTANLNRPAKRYFPPTTTGNAVASGMMDAVCGSIMM	390

An ORF61ng nucleotide sequence <SEQ ID 237> was predicted to encode a protein having amino acid sequence <SEQ ID 238>:

```

1  MFSFGWAFDR PQYELGSLSP VAALACRRAL GCLGLETQIK WPNDLVVGRD
5  51  KLGILIIETV RAGGKTVAUV GIGINFVLPK EVENAASVQS LFQTASRRGN
101 ADAAVLLET L LAELGAVLEQ YAEEGFAPFL NEYETANRDH GKAVLLLRDG
151 ETVCEGTVKG VDGRGVHLHLE TAEGEQTVVS GEISLRPDNR SVSVPKRPDS
201 ERFLLLLEGGN SRLKWAVVEN GTFATVGSAP YRDLSPLGAE WAEKADGNVR
251 IVGCAVCGES KKAQVKEQLA RKIEWLPSSA QALGIRNHYR HPEEHGSDRW
301 FNALGSRRFS RNACVVVSCG TAVTVDALTD DGHYLGGTIM PGFHLMKESL
10 351 AVRTANLNRP AGKRYPFPTT TGNASVSGMM DAVCGSIMMM HGRLKEKNGA
401 GKPVDVIITG GGAAKVAEAL PPAFLAENTV RVADNLVIHG LLNLIAAEGG
451 ESEHA*

```

Further analysis revealed the complete gonococcal DNA sequence <SEQ ID 239> to be:

```

1  ATGACGGTTT TGAAGCCTTC GCATTGGCGG GTGTTGGCGG AGCTTGCCGA
15 51  CGGTTTGCGG CAACACGTAT CGCAATTGGC GCGTGAGGCG GACATGAAGC
101 CGCAGCAGCT CAACGGTTTT TGGCAGCAGA TGCCGGCGCA TATACGCGGG
151 CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CTTTGGCGGT
201 TTTGATGCC GAAGGTTTGC GCGATCTGGG GGAAAGGTCG GGTTCCTAGA
251 GGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
20 301 GCGCGGATTG CGCCGGACAA GGCGCACAAA ACCATATGCG TGACCCACCT
351 GCAAAGTAAG GGCAGGGGGC GGCAGGGGCG GAAGTGGTCG CACCGTTTGG
401 GCGAGTGCCT GATGTTCACT TTCGGCTGGG CGTTTGACCG GCCGCGAGTAT
451 GAGTTGGGTT CGCTGTGCGC TGTGCGGCA CTGCGTGCC GCGCGCGTTT
501 GGGGTGTTTG GGTGTTGAAA CGCAATCAA GTGGCCAAAC GATTGTTGTCG
25 551 TCGGACGCGA CAAATTGGGC GGCATTCTGA TTGAAACAGT CAGGGCGGGC
601 GGTAACACGG TTGCCGTGGT CCGTATCGGC ATCAATTTTCG TGCTGCCCAA
651 GGAAGTGAA AACGCCGCTT CCGTGAGTC GCTGTTTCAG ACGGCATCGC
701 CGAGCGGGCA TGCCGATGCC GCCGTATTGC TGGAAACATT GCTTGCGGAA
751 CTGGGCGCGG TGTGGAACA ATATGCGGAA GAAGGGTTCG CGCCATTTTT
30 801 AAATAGATAT GAAACGGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT
851 TGCGCGACGG CGAAACCGTG TGCGAAGGCA CGGTTAAAGG CGTGGACGGA
901 CGAGCGGTTT TGCATTGGA AACGGCAgaa ggcgaACaGa cggtcgtcag
951 cggcgaaaTC AGcctGCggc cgcacaacaG GTCGGtttcc gtgcccgaagc
1001 ggccggatTC GgaacgtTTT tTGctgttgg aaggcgggaa cagccgGCTC
35 1051 AAGTGGGCGT GggtggAAaA cggcacgttc gcaaccgtgg gcagcgcgCc
1101 gtaCCGCGAT TTGTCGCCCT TGGGCGCGGA GTGGGCGGAA AAGGCGGATG
1151 GAAATGTCCG CATCGTCGGT TGGCGCGTGT GCGGAGAATC CAAAAAGGCA
1201 CAAGTGAAGG AACAGCTCGC CCGAAAATC GAGTGGCTGC CGTCTTCCGC
1251 ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCCGAA GAACACGGTT
40 1301 CAGACCGTTG GTTCAACGCC TTGGGCGAGC GCCGCTTCAG CCGCAACGCG
1351 TGCGTCGTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCGA
1401 TGACGGACAT TATCTCGGCG GAACCATCAT GCCCGGCTTC CACCTGATGA
1451 AAGAATCGCT CGCCGTCCGA ACCGCCAACC TCAACCGCCC CGCCGGCAAA
1501 CGTTACCCTT TCCCAGCAC AACGGGCAAC GCCGTCGCAA GCGGCATGAT
45 1551 GGACGCGGTT TCGCGCTCGA TAATGATGAT GCACGGCCGT TTGAAAGAAA
1601 AAAACGGCGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CGGCGGCGCG
1651 GCGAAAGTCG CCGAAGCCCT GCCGCTGCA TTTTGGCGG AAAATACCGT
1701 GCGCGTGGCG GACAACCTCG TCATCCACGG GCTGCTGAAC CTGATTGCCG
1751 CCGAAGCGCG GGAATCGGAA CACGCTTAA

```

50 This corresponds to the amino acid sequence <SEQ ID 240; ORF61ng-1>:

```

1  MTVLKPSHWR VLAELADGLP QHVSQAREA DMKPQQLNGF WQMPAHIRG
51  LLRQHDGYWR LVRPLAVFDA EGLRDLGERS GFQTALKHEC ASSNDEILEL
101 ARIAPDKAHK TICVTHLQSK GRGRQGRKWS HRLGECIMFS FGWAFDRPOY
55 151 ELGSLSPVAA LACRRALGCL GLETQIKWPN DLVVGRDKLG GILIIETVRAG
201 GKTVAUVGIG INFVLPKEVE NAASVQSLFQ TASRRGNADA AVLLETLLAE
251 LGAVLEQYAE EGFAPFLNEY ETANRDHGKA VLLLRDGETV CEGTVKGVVDG
301 RGVHLHLETA GEQTVVSGEI SLRPDNRVS VPKRPDSERF LLLEGGNSRL
351 KWAVWENGTF ATVGSAPYRD LSPLGAEWAE KADGNVRIVG CAVCGESKKA
401 QVKEQLARKI EWLPSAQAAL GIRNHRYRHE EHGSDRWANA LGSRRFSRNA
60 451 CVVVSCGTAV TVDALTDGHH YLGGTIMPGF HLMKESLAVR TANLNRPAGK
501 RYFPFTTGN AVASGMMDAV CGSIMMHGR LKEKNGAGKP VDVIIITGGGA
551 AKVAEALPPA FLAENTVRVA DNLVIHGLLN LIAAEGGESE HA*

```

ORF61ng-1 and ORF61-1 show 93.9% identity in 591 aa overlap:

	orf61ng-1.pep	MTVLKPSHWRVLAELADGLPQHVSQALAREADMKPQQLNGFWQOMPAHIRGILLRQHDGYWR	60
	orf61-1	MTVLKLSHWRVLAELADGLPQHVSQALARMADMKPQQLNGFWQOMPAHIRGILLRQHDGYWR	60
5	orf61ng-1.pep	LVRPLAVFDAEGLRDLGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK	120
	orf61-1	LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK	120
10	orf61ng-1.pep	GRGRQGRKWSHRLGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPN	180
	orf61-1	GRGRQGRKWSHRLGECLMFSFGWVDFRPQYELGSLSPVAACRRALSRLGLDVQIKWPN	180
15	orf61ng-1.pep	DLVVGRDKLGGILITVTRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA	240
	orf61-1	DLVVGRDKLGGILITVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA	240
20	orf61ng-1.pep	AVLLETLLAELGAVLEQYAEFGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVVDG	300
	orf61-1	AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVVDG	300
25	orf61ng-1.pep	RGVLHLETAEGEQTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGNSRLKWAWVENGTG	360
	orf61-1	QGVLLHLETAEGKQTVVSGEISLRSDRPVSVPKRRDSERFLLLDGNSRLKWAWVENGTG	360
30	orf61ng-1.pep	ATVGSAPYRDLSPGLAEWAEEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQAL	420
	orf61-1	ATVGSAPYRDLSPGLAEWAEEKADGNVRIVGCAVCGEFKKAQVQEQQLARKIEWLPSSAQAL	420
35	orf61ng-1.pep	GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDGHHYLGGTIMPGF	480
	orf61-1	GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDGHHYLGGTIMPGF	480
40	orf61ng-1.pep	HLMKESLAVRTANLNRPAKRYFPFPTTGNASGMMDAVCGSIMMHGRLKEKNGAGKP	540
	orf61-1	HLMKESLAVRTANLNRHAGKRYFPFPTTGNASGMMDAVCGSVMMHGRLEKKTGAGKP	540
	orf61ng-1.pep	VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNIAAEGGESEHAX	593
	orf61-1	VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNIAAEGREYEHIX	593

Based on this analysis, including the homology with the baf protein of *B.pertussis* and the presence of a putative prokaryotic membrane lipoprotein lipid attachment site, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### 45 Example 29

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 241>:

	1	ATGTTTACC	AAATCCTTGC	CCTGATTATC	TGGAGCAGCT	CGTTTATTGC
	51	CGCCAAATAT	GTCTATGGCG	GCATCGATCC	CGCATTGATG	GTCGGCGTGC
50	101	GCCTGCTAAT	TGCCGCGCTG	CCTGCACTGC	CCGCTGCCG	CCGTGATGTC
	151	GGCAAGATTC	CGCGTGAGGA	ATGGAAGCCG	TTGCTGATTG	TGTCGTTTCTG
	201	CAACTATGTG	CTGACCCTGC	TGCTTCAGTT	TGTCGGGTTG	AAATACACTT
	251	CGCCCGCCAG	CGCATCGGTC	ATTGTCGGAC	TCGAGCCGCT	GCTGATGGTG
	301	TTGTGCGGAC	ACTTTTCTT	CAACGACAAA	GCGCGTGCCT	ACCACTGGAT
	351	ATGCGGCGCG	GCGGCATTTC	CCGGTGTCGC	GCTGCTGATG	GCGGGCGGTG
55	401	CGGAGAGGG	CGGCGAAGTC	GGCTGGTTCG	GCTGCCTGCT	GGTGTGTTG
	451	GCGGGCGCGG	GCTTTTGTGC	CGCTATGCGT	CCGACGCAAA	GGCTGATTGC
	501	ACGCATCGGC	GCACCGCAT	TCACATCTGT	TTCCATTGCC	GCCGCATCGT
	551	TGATGTGCCT	GCGTTTTCG	CTTGCTTTGG	CGCAAAGTTA	TACCGTGGAC
	601	TGGAGCGTCG	GGATGTATT	GTGCTGCTG	TATTTGGGTT	TGGGGTGC..

60 This corresponds to the amino acid sequence <SEQ ID 242; ORF62>:



-177-

1 MFYQILALII WSSSFIAAKY VYGGIDPALM VGVRLIIAAL PALPACRRHV  
 51 GKIPREEWKP LLIVSFVNYV LTLQLQFVGL KYTSAASASV IVGLEPLLMV  
 101 FVGHHFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL  
 151 AGAGFCAAMR PTORLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD  
 5 201 WSVGMVLSLL YLGLGC..

Further work revealed the complete nucleotide sequence <SEQ ID 243>:

1 ATGTTTTACC AAATCCTTGC CCTGATTATC TGGAGCAGCT CGTTTTATTGC  
 51 CGCCAAATAT GTCTATGGCG GCATCGATCC CGCATTGATG GTCGGCGTGC  
 101 GCCTGCTAAT TGCCGCGCTG CCTGCACTGC CCGCCTGCCG CCGTCATGTC  
 151 GGCAAGATTC CGCGTGAGGA ATGGAAGCCG TTGCTGATTG TGTGCTTCGT  
 201 CAACTATGTG CTGACCCTGC TGCTTCAGTT TGTGCGGTTG AAATACACTT  
 251 CCGCCGCCAG CGCATCGGTC ATTGTCCGAC TCGAGCCGCT GCTGATGGTG  
 301 TTTGTCGGAC ACTTTTTCTT CAACGACAAA GCGCGTGCCT ACCACTGGAT  
 351 ATGCGGCGCG CGGCGATTG CCGGTGTGCG GCTGCTGATG GCGGCGGTTG  
 15 401 CGGAAGAGGG CGGCGAAGTC GGCTGGTTCG GCTGCCTGCT GGTGTTGTTG  
 451 GCGGCGCGCG GCTTTTGTGC CGCTATGCGT CCGACGCAAA GGCTGATTGC  
 501 ACGCATCGGC GCACCGGCAT TCACATCTGT TTCCATTGCC GCCGCATCGT  
 551 TGATGTGCCT GCCGTTTTCG CTTGCTTTGG CGCAAAGTTA TACCGTGGAC  
 601 TGGAGCGTCG GGATGGTATT GTCGCTGCTG TATTGGGTT TGGGGTGCGG  
 20 651 CTGGTACGCC TATTGGCTGT GGAACAAGGG GATGAGCCGT GTTCTGCCA  
 701 ATGTTTCGGG ACTGTTGATT TCGCTCGAAC CCGTCGTCGG CGTGCTGCTG  
 751 GCGGTTTTGA TTTTGGGCGA ACACCTGTCG CCCGTGTCCG CCTTGGGCGT  
 801 GTTTGTCGTC ATCGCCGCCA CCTTGGTTGC CGGCCGGCTG TCGCATCAAA  
 851 AATAA

25 This corresponds to the amino acid sequence <SEQ ID 244; ORF62-1>:

1 MFYQILALII WSSSFIAAKY VYGGIDPALM VGVRLIIAAL PALPACRRHV  
 51 GKIPREEWKP LLIVSFVNYV LTLQLQFVGL KYTSAASASV IVGLEPLLMV  
 101 FVGHHFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL  
 151 AGAGFCAAMR PTORLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD  
 30 201 WSVGMVLSLL YLGLGCGWYA YWLWNKMSR VPANVSGLLI SLEPVVGVL  
 251 AVLILGEHLS PVSALGVFVV IAATLVAGRL SHQK\*

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical transmembrane protein HI0976 of *H. influenzae* (accession number Q57147)

ORF62 and HI0976 show 50% aa identity in 114aa overlap:

35 Orf62 1 MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRRXXXXXXXXXXCRRHVGKIPREEWKP 60  
 M YQILAL+IWSSS I K Y +DP L+V VR R KI + K  
 HI0976 1 MLYQILALLIWSSSLIVGKLTYSMDPVLVVQVRLIIAMIIVMPLFLRRWKKIDKPMRKQ 60  
 40 Orf62 61 LLIVSFVNYVLTLLQLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAY 114  
 L ++F NY LLQF+GLKYTSA+SA ++GLEPLL+VFVGHFFF K +  
 HI0976 61 LWWLAFFNYTAVFLQLQFVGLKYTSAASAVTMIGLEPLLVVVFVGHFFFKTKQNGF 114

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF62 shows 99.5% identity over a 216aa overlap with an ORF (ORF62a) from strain A of *N.*

45 *meningitidis*:

10 20 30 40 50 60  
 orf62.pep MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHVGKIPREEWKP  
 orf62a MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHVGKIPREEWKP  
 50 10 20 30 40 50 60  
 70 80 90 100 110 120  
 orf62.pep LLIVSFVNYVLTLLQLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA  
 orf62a LLIVSFVNYVLTLLQLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA  
 55 70 80 90 100 110 120  
 130 140 150 160 170 180  
 orf62.pep AAFAGVALLMAGGAEEGGEVWFGCLLVLLAGAGFCAAMRPTORLIARIGAPAFTSVSIA

```

|||||
orf62a      AAFAGVALLMAGGAEEGGEVGFFGCLLVLLAGAGFCAAMRPTORLIRIAGAPAFTSVSIA
              130      140      150      160      170      180

5
              190      200      210
orf62.pep    AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGC
|||||
orf62a      AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGVGCSWYAYWLWNKGMSRVPANVSGLLI
              190      200      210      220      230      240

10
orf62a      SLEPVVGVLLAVLILGEHLSSPVSVLGVFVVIAATLVAGRLSHQKX
              250      260      270      280

```

The complete length ORF62a nucleotide sequence <SEQ ID 245> is:

15	1	ATGTTTACC	AAATCCTTGC	CCTGATTATC	TGGAGCAGCT	CGTTTATTGC
	51	CGCCAAATAT	GTCTATGGCG	GCATCGATCC	CGCATTGATG	GTCCGGCGTGC
	101	GCCTGCTGAT	TGCTGCGCTG	CCTGCACTGC	CCGCGTCCCG	CCGTCATGTC
	151	GGCAAGATTG	CGCGTGAGGA	ATGGAAGCCG	TTGCTGATTG	TGTCGTTCCG
	201	CCACTATGTG	CTGACCCTGC	TACTTCAGTT	TGTCCGGTTG	AAATACACTT
20	251	CCGCCGCCAG	CGCATCGGTC	ATTGTCGGAC	TCGAGCCACT	GCTGATGGTG
	301	TTTGTGCGGAC	ACTTTTTCTT	CAACGACAAA	GCGCGTGCGT	ACCACTGGAT
	351	ATGCGGCGCG	CGGGCATTTC	CCGGTGTCGC	GCTGCTGATG	GCGGGCGGGTG
	401	CGGAAGAGGG	CGCGAAGTC	GGCTGCTTCG	GCTGCTGCTG	GGGTGTGTTG
	451	GCGGGCGCGG	GCTTTTGTGC	CGCTATGCGT	CCGACGCAAA	GGCTGATTGC
25	501	ACGCATCGGC	GCACCGGCAT	TCACATCTGT	TTCCATTGCC	GCCGCATCGT
	551	TGATGTGCCT	GCCGTTTTTCG	CTTGCTTTGG	CGCAAAGTTA	TACCGTGGAC
	601	TGGAGCGTCG	GAATGGTATT	GTCGCTGCTG	TATTTGGGCG	TGGGGTGCAG
	651	CTGGTACGCC	TATTGGCTGT	GGAACAAGGG	GATGAGCCGT	GTTCCTGCCA
	701	ACGTTTCGGG	ACTGTTGATT	TCGCTCGAAC	CCGTCGTCGG	CGTGCTGCTG
30	751	GCGGTTTTGA	TTTTGGGCGA	ACACTGTTCG	CCCGTGTCGG	TCCTGGGCGT
	801	GTTTGTGCTC	ATCGCCGCCA	CCTTGGTTGC	CGGCCGGCTG	TCGCATCAAA
	851	AATAA				

**This encodes a protein having amino acid sequence <SEQ ID 246>:**

35

1	<u>MFYQILALII</u>	<u>WSSSFIAAKY</u>	<u>VYGGIDPALM</u>	<u>VGVRLLIAAL</u>	<u>PALPACRRHV</u>
51	<u>GKIPREEWKP</u>	<u>LLIVSFVNYV</u>	<u>LTLLQLQFVGL</u>	<u>KYTSAAASASV</u>	<u>IVGLEPLLMV</u>
101	<u>VGVGHFFNDK</u>	<u>ARAYHWICWA</u>	<u>AAFAGVALLM</u>	<u>AGGAEGGEVW</u>	<u>GWFGCLLVLL</u>
151	<u>AGAGFCAAMR</u>	<u>PTQRLIARIG</u>	<u>APAFTSVSIA</u>	<u>AASLMCLPFS</u>	<u>LALAQSYTVD</u>
201	<u>WSVGMVLSLL</u>	<u>YLGWCSGYSA</u>	<u>YWLWNKGMSR</u>	<u>VPANVSGLLI</u>	<u>SLEPVGVLL</u>
251	<u>AVLILGEHLS</u>	<u>PVSVLGVEVV</u>	<u>IATLTVAGRL</u>	<u>SHOK*</u>	

**ORF62a and ORF62-1 show 98.9% identity in 284 aa overlap:**

40	orf62a.pep	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHVKGIPREEWKP	60
	orf62-1	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHVKGIPREEWKP	60
45	orf62a.pep	LLIVSFVNYVLTLLQLFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA	120
	orf62-1	LLIVSFVNYVLTLLQLFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA	120
50	orf62a.pep	AAFAGVALLMAGGAEEGGEVGVFGCLLVLLAGAGFCAAMRPTQRLIARIGAPFTSVSIA	180
	orf62-1	AAFAGVALLMAGGAEEGGEVGVFGCLLVLLAGAGFCAAMRPTQRLIARIGAPFTSVSIA	180
55	orf62a.pep	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGVGCSWYAYWLWNKGMSRVPANVSGLLI	240
	orf62-1	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANVSGLLI	240
	orf62a.pep	SLEPVVGVLLAVLILGEHLSPVSVLGVFVVIATLVAGRLSHQKX	285
	orf62-1	SLEPVVGVLLAVLILGEHLSPVSVLGVFVVIATLVAGRLSHQKX	285

60 Homology with a predicted ORF from *N.gonorrhoeae*

ORF62 shows 99.5% identity over a 216aa overlap with a predicted ORF (ORF62.ng) from *N. gonorrhoeae*:

	orf62.pep	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGVKIPREEWKP	60
	orf62ng	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGVKIPREEWKP	60
5	orf62.pep	LLIVSFVNYVLTLLQLQFVGLKYTSAASASVIVGLEPLLMVFGVGHFFNDKARAYHWICGA	120
	orf62ng	LLIVSFVNYVLTLLQLQFVGLKYTSAASASVIVGLEPLLMVFGVGHFFNDKARAYHWICGA	120
10	orf62.pep	AAFAGVALLMAGGAEEGGEVWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFSTSVSIA	180
	orf62ng	AAFAGVALLMAGGAEEGGEVWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFSTSVSIA	180
	orf62.pep	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGC	216
15	orf62ng	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANASGLLI	240

The complete length ORF62ng nucleotide sequence <SEQ ID 247> is:

	1	ATGTTTACC	AAATCCTTGC	CCTGATTATC	TGGGGCAGCT	CGTTTATTGC
	51	CGCCAAATAT	GTCTATGGCG	GCATCGATCC	CGCATTGATG	GTCCGGCGTGC
20	101	GCCTGCTGAT	TGCCGGCGTG	CCTGCACTGC	CCGCCTGCCG	CCGTCATGTC
	151	GGCAAGATTC	CGCGTGAGGA	ATGGAAGCCG	TTGCTGATTG	TGTCGTTCTG
	201	CAACTATGTG	CTGACCTGCG	TGCTTCAGTT	TGTCGGGTTG	AAATACACTT
	251	CCGCCGCCAG	CGCATCCGTC	ATTGTCGGAC	TCGAGCCGCT	GCTGATGGTG
	301	TTTGTCCGAC	ACTTTTCTT	CAACGACAAA	GCGCGTGCC	ACCACTGGAT
25	351	ATGCGGCGCG	GCGGCATTTC	CCGGTGTCGC	GCTGCTGATG	GCGGGCGGTG
	401	CGGAAGAGGG	CGGCGAAGTC	GGCTGGTTCG	GCTGCCTGCT	GGTGTGTTG
	451	GCGGGCGCGG	GCTTTTGTGC	CGCTATGCGT	CCGACGCAAA	GGCTGATTGC
	501	CCGCATCGGC	GCACCGGCAT	TCACATCTGT	TTCCATTGCC	GCCGCATCGT
	551	TGATGTGCCT	GCCGTTTTCG	CTTGCTTTGG	CGCAAAGTTA	TACCGTGGAC
30	601	TGGAGCGTCG	GGATGTTATT	GTCGCTGTTG	TATTTGGGTT	TGGGGTGC
	651	CTGGTACGCC	TATTGGCTGT	GGAAACAAGG	GATGAGCCGT	GTTCTGCGCA
	701	ACGCGTCGGG	ACTGTTGATT	TCGCTCGAAC	CCGTCGTCGG	CGTGCTGTTG
	751	GCGGTTTTGA	TTTTGGGCGA	ACATTTATCG	CCCGTGTCCG	CCTTGGGCGT
	801	GTTTGTGCTC	ATCGCCGCCA	CTTTCGCCGC	CGGCCGGCTG	TCGCGCAGGG
	851	ACCGCAAAA	CGGCAATGCC	GTCTGA		

35 This encodes a protein having amino acid sequence <SEQ ID 248>:

	1	<u>MFYQILALII</u>	<u>WGSSFIAAKY</u>	<u>VYGGIDPALM</u>	<u>VGVRLLIAAL</u>	<u>PALPACRRHV</u>
	51	<u>GKIPREEWKP</u>	<u>LLIVSFVNYV</u>	<u>LTLLQLQFVGL</u>	<u>KYTSAAASV</u>	<u>IVGLEPLLMV</u>
	101	<u>FVGHFFNDK</u>	<u>ARAYHWICGA</u>	<u>AAFAGVALLM</u>	<u>AGGAEEGGEV</u>	<u>GWFGCLLVLL</u>
40	151	<u>AGAGFCAAMR</u>	<u>PTORLIARIG</u>	<u>APAFSTSVSIA</u>	<u>AASLMCLPFS</u>	<u>LALAQSYTVD</u>
	201	<u>WSVGMVLSLL</u>	<u>YLGLGCGWYA</u>	<u>YWLWNKGMSR</u>	<u>VPANASGLLI</u>	<u>SLEPVGVLL</u>
	251	<u>AVLILGEHLS</u>	<u>PVSALGVFVV</u>	<u>IAATFAAGRL</u>	<u>SRRDAQNGNA</u>	<u>V*</u>

ORF62ng and ORF62-1 show 97.9% identity in 283 aa overlap:

		10	20	30	40	50	60
45	orf62ng.pep	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGVKIPREEWKP					
	orf62-1	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGVKIPREEWKP					
		10	20	30	40	50	60
50	orf62ng.pep	LLIVSFVNYVLTLLQLQFVGLKYTSAASASVIVGLEPLLMVFGVGHFFNDKARAYHWICGA					
	orf62-1	LLIVSFVNYVLTLLQLQFVGLKYTSAASASVIVGLEPLLMVFGVGHFFNDKARAYHWICGA					
		70	80	90	100	110	120
55	orf62ng.pep	AAFAGVALLMAGGAEEGGEVWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFSTSVSIA					
	orf62-1	AAFAGVALLMAGGAEEGGEVWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFSTSVSIA					
		130	140	150	160	170	180
60	orf62ng.pep	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANASGLLI					
	orf62-1	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANASGLLI					
		190	200	210	220	230	240
65	orf62ng.pep	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANASGLLI					
	orf62-1	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANASGLLI					
		190	200	210	220	230	240

-180-

```

                250      260      270      280      290
orf62ng.pep    SLEPVVGVLLAVLILGEHLSPVSALGVFVVIATFAAGRLSRRDAQNGNAVX
                |||||
5  orf62-1      SLEPVVGVLLAVLILGEHLSPVSALGVFVVIATLVAGRLSHQKX
                250      260      270      280

```

Furthermore, ORF62ng shows significant homology to a hypothetical *H. influenzae* protein:

```

10  sp|Q57147|Y976_HAEIN HYPOTHETICAL PROTEIN HI0976 >gi|1074589|pir||B64163
    hypothetical protein HI0976 - Haemophilus influenzae (strain Rd KW20)
    >gi|1574004 (U32778) hypothetical [Haemophilus influenzae] Length = 128
    Score = 106 bits (262), Expect = 2e-22
    Identities = 56/114 (49%), Positives = 68/114 (59%)

15  Query: 1   MFYQILALIWGSSFIAAKYVYGGIDPALMVGVRRXXXXXXXXXXXXRRHVKGIPREWKP 60
    M YQILAL+IW SS I K Y +DP L+V VR R KI + K
    Sbjct: 1   MLYQILALLIWSSSLIVGKLTYSMMDPVLVVQVRLIAMIIVMPLFLRRWKIKDKPMRKQ 60

    Query: 61  LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLVMFVGHHFFNDKARAY 114
    L ++F NY LLQF+GLKYTSA+SA ++GLEPLL+VFVGHHFF K +
20  Sbjct: 61  LWWLAFFNYTAVFLLQFVGLKYTSAASAVTMIGLEPLLVMFVGHHFFKTKQNGF 114

```

Based on this analysis, including the homology with the transmembrane protein of *H. influenzae* and the putative leader sequence and several transmembrane domains in the gonococcal protein, it is predicted that these proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 30

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 249>:

```

30  1  ATGCGCCGTT TTCTACCGAT CGCAGCCATA TGCGCmGwms TCCTGkkGTA
    51  sGGACTGACG GCGGCAACCG GCAGCACCAG TTCGCTGGCG GATTATTTCT
    101 GGTGGATTGT TCGGTTGAGC GCAATGCTGC TGCTGGTGTG GTCCGCCGTT
    151 TTGGCACGTT ATGTCATATT GCTGTTGAAA GACAGGCGCG ACGGCGTATT
    201 CGGTTTCGCTA srTyGCCAAA gsGCCTgkks TGGG.ATGTT TACGCTGGTT
    251 GCGGKACTGC CCGGCGTGTG TCTGTTGCGC TTCCCGCAC AGTTCATCAA
35  301 CGGCACGATT AATTCGTGGT TCGGCAACGA TACCCACGAG GCGCTGAAC
    351 GCAGCCTCAA TTTGAGCAAG TCCGCATTGA ATTTGGCGGC AGACAACGCC
    401 CTCGGCAACG CCGTCCCCGT GCAGATAGAC CTCATCGGCG CGGCTTCCCT
    451 GCCCGGGGAT ATGGGCAGGG TGCTGGAACA TTACGCCGCG AGCGGTTTGT
    501 CCCAGCTTGC CCTGTACAAY ksCGCAAGCG GCAAATCGA AAAAAGCATC
    551 AACC CGCACA AGCTCGATCA GCCGTTTCCA GGTAAGGCGC GTTGGGAaAa
40  601 AATCCaACGG GCGGGTTCGG TCAGGGATTT GGAAAGCATA GCGGCGGTAT
    651 TGTaCGCGCA GGGCTGGCTG TCGGCGGGTA CGCACwACGG GCGCGATTAC
    701 GCCTTGTTTT TCGCTCAGCC GGTTCCTCAA GGCGTGGCAG AGGATGCCGT
    751 yTTAATCGAA AAGGCAAGGG CGAAATATGC TGAGTTGAGT TACAGCAAAA
    801 AAGGTTTGCA GACCTTTTTC CTGGCAACCC TGCTGATTGC CTCGCTGCTG
45  851 TCGATTTTTC TTGCACTGGT CATGGCACTG TATTCGCCC GCCGTTTCGT
    901 CGAACCCGTC CTATCGCTTG CCGAGGGGGC GAAGGCGGTG GCGCAAGGCG
    951 ATTCAGCCA GACGCGCCCC GTGTTGCGCA ACGACGAGTT CGGACGCTTG
    1001 ACCaRGTTGT TCAACCACAT GACCAGCAG CTTTCCATCG CCAAAGATGC
    1051 AGACGAGCGC AACC GCGGCG GCGAGGAAGC CGCCAGGCAT TATCTTGAAT
50  1101 CCGTGTGTTGA GGGGCTGACC ACGGCGGTGG TGGTGTGTTGA CGAACAAGGC
    1151 TGTCTGAAAA CCTTCAACAA AGCGCGGGT ACC..

```

This corresponds to the amino acid sequence <SEQ ID 250; ORF64>:

```

55  1  MRRFLPIAAI CAXXLXXGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV
    51  LARYVILLK DRRDGVFGSX XAKXPXXMF TLVAXLPGVF LFGFPAQFIN
    101 GTINSWFGND THEALERSLN LSKSALNLAA DNALGNAVVP QIDLIGAASL
    151 PGDMGRVLEH YAGSGFAQLA LYNXASGKIE KSINPHKLDQ PFPKGARWEK
    201 IQRAGSVRDL ESIGGVLYAQ GWLSAGTHXG RDYALFFRQP VPKGVAEDAV
    251 LIEKARAKYA ELSYSKKGLQ TFFLATLLIA SLLSIFLALV MALYFARRFV

```

301 EPVLSLAEGA KAVAQGDfsQ TRPVLRNDEF GRLTXLFNHM TEQLSIAKDA  
 351 DERNRRREEA ARHYLECVLE GLTTGVVVFDE EQGCLKTFNK AAGT..

Further work revealed the complete nucleotide sequence <SEQ ID 251>:

```

      1 ATGCGCCGTT TTCTACCGAT CGCAGCCATA TCGCGCGTCG TCCTGTTGTA
5      51 CGGACTGACG GCGGCAACCG GCAGCACCAG TTCGCTGGCG GATTATTTCT
     101 GGTGGATTGT TCGGTTTCAGC GCAATGCTGC TGCTGGTGTT GTCCGCCGTT
     151 TTGGCACGTT ATGTCATATT GCTGTTGAAA GACAGGCGCG ACGGCGTATT
     201 CGGTTTCGCAG ATTGCCAAAC GCCTTTCTGG GATGTTTACG CTGGTTGCCG
     251 TACTGCCCGG CGTGTCTCTG TTCGGCGTTT CCGCACAGTT CATCAACGGC
    10      301 ACGATTAAAT CGTGGTTCGG CAACGATACC CACGAGGCGC TTGAACGCAG
     351 CCTCAATTG AGCAAGTCCG CATTGAATTT GGCGGCAGAC AACGCCCTCG
     401 GCAACGCCGT CCCCGTCAG ATAGACCTCA TCGGCGCGGC TTCCTGCCCC
     451 GGGGATATGG CAGGGTGCT GGAACATTAC GCCGGCAGCG GTTTTGCCCA
     501 CGTTGCCCTG TACATGCCG CAAGCGGCAA AATCGAAAAA AGCATCAACC
    15      551 CGCACAAGCT CGATCAGCCG TTTCCAGGTA AGGCGCGTTG GAAAAAATC
     601 CAACGGGCGG GTTCGGTCAG GGATTTGGAA AGCATAGGCG GCGTATTGTA
     651 CGCGCAGGGC TGGCTGTCTG CGGGTACGCA CAACGGGCGC GATTACGCTT
     701 TGTTTTTCCG TCAGCCGGTT CCCAAAGGCG TGGCAGAGGA TGCCGTCTTA
     751 ATCGAAAAGG CAAGGGCGAA ATATGCTGAG TTGAGTTACA GCAAAAAGG
    20      801 TTTGCAGACC TTTTCTCTGG CAACCCTGCT GATTGCCTCG CTGCTGTCTGA
     851 TTTTCTTGC ACTGGTCATG GCACTGTATT TCGCCCGCCG TTTGCTCGAA
     901 CCCGTCTTAT CGCTTGCCGA GGGGGCGAAG GCGGTGCGCG AAGGCGATTT
     951 CAGCCAGACG CGCCCGTGT TCGCAACGCA CGAGTTCGGA CGCTTGACCA
    25     1001 AGTTGTTCAA CCACATGACC GAGCAGCTTT CCATCGCCAA AGAAGCAGAC
     1051 GAGCGCAACC GCCGGCGCGA GGAAGCCGCC AGGCATTATC TTGAATGCGT
     1101 GTTGAGGGGG CTGACCACGG CCGTGGTGGT GTTTGACGAA CAAGGCTGTC
     1151 TGAAAACCTT CAACAAAGCG GCGGAACAGA TTTTGGGGAT GCCGCTTACC
     1201 CCCCTGTGGG GCAGCAGCCG GCACGGTTGG CACGGCGTTT CGGCGCAGCA
     1251 GTCCCTGCTT GCCGAAGTGT TTGCCGCCAT CGGCGCGCGC GCAGGTACGG
    30     1301 ACAAAACCGG CCATGTGAAA TATGCCGCGC CGGACGATGC CAAAATCCTG
     1351 CTGGGCAAGG CAACCGTCTT GCCGAAGAC AACGGCAACG GCGTGGTAAT
     1401 GGTGATTGAC GACATCACCG TTTTGATACA CGCGCAAAAA GAAGCCGCGT
     1451 GGGGCGAAGT GGCGAAGCGG CTGGCACACG AAATCCGCAA TCCGCTCACG
     1501 CCCATCCAGC TTTCCGCCGA ACGGCTGGCG TGGAATTTGG GCGGGAAGCT
    35     1551 GGATGAGCAG GATGCCGAAA TCCTGACGCG TTCGACCGAC ACCATCGTCA
     1601 AACAGGTGGC GGCATTGAAG GAAATGGTCG AAGCATTCGG CAATTATGCG
     1651 CGTTCCCTT CGCTCAAATT GGAAATCAG GATTTGAACG CCTTAATCGG
     1701 CGATGTGTTG GCATTGTATG AAGCCGGTCC GTGCCGGTTT GCGGCGGAGC
     1751 TTGCCGGCGA ACCGCTGACG GTGGCGGCGG ATACGACCGC CATGCGGCAG
    40     1801 TTGCTGCACA ATATTTTCAA AAATGCCGCC GAAGCGGCGG AAGAAGCCGA
     1851 TGTGCCCGAA GTCAGGGTAA AATCGGAAAC AGGGCAGGAC GGTCCGATTG
     1901 TCCTGACGGT TTGCGACAAC GGCAAGGGT TCGGCAGGGA AATGCTGCAC
     1951 AACGCCTTCG AGCCGTATGT AACGGACAAA CCGGCGGGAA CGGGATTGGG
     2001 TCTGCCTGTG GTGAAAAAAA TCATTGAAGA ACACGGCGGC CGCATCAGCC
    45     2051 TGAGCAATCA GGATGCGGGT GGCGCGTGTG TCAGAATCAT CTTGCCAAAA
     2101 ACGGTAAAAA CTTATGCGTA G
  
```

This corresponds to the amino acid sequence <SEQ ID 252; ORF64-1>:

```

      1 MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVAFS AMLLLVL SAV
    50     51 LARYVILLK DRDGVFGSQ IAKRLSGMET LVAVLPGVFL FGVSAQFING
     101 TINSWFGNDT HEALERSLNL SKSALNLAAD NALGNAVVPVQ IDLIGAASLP
     151 GDMGRVLEHY AGSGFAQLAL YNAASGKIEK SINPHKLDQP FPGKARWEKI
     201 QRAGSVRDLE SIGGVLYAQG WLSAGTHNGR DYALFFRQPV PKGVAEDAVL
     251 IEKARAKYAE LSYSKKGLQT FFLATLLIAS LLSIFLALVM ALYFARRFVE
    55     301 PVLSLAEGAK AVAQGDfsQT RPVLRNDEF GRLTXLFNHMT EQLSIAKEAD
     351 ERNRRREEAA RHYLECVLEG LTTGVVVFDE EQGCLKTFNK AAEQILGMPLT
     401 PLWGSRRHWG HGVSAQSSLL AEFVFAAIGAA AGTDKPVHVK YAAPDDAKIL
     451 LGKATVLPED NGNGVVMVID DITVLIHAQK EAAWGEVAKR LAHEIRNPLT
     501 PIQLSAERLA WKLGGKLDEQ DAQILTRSTD TIVKQVAALK EMVEAFRNYA
     551 RSPSLKLENQ DLNALIGDVL ALYEAGPCRF AAELAGEPLT VAADTTAMRQ
    60     601 VLHNIFKNAA EAAEEADVPE VRVKSETGQD GRIVLTVCDN GKGFREMLH
     651 NAFEPYVTDK PAGTGLGLPV VKKIIIEHGG RISLSNQDAG GACVRIILPK
     701 TVKTYA*
  
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF64 shows 92.6% identity over a 392aa overlap with an ORF (ORF64a) from strain A of *N. meningitidis*:

5	orf64.pep	10 20 30 40 50 60 MRRFLPIAAICAXXLXGLTAATGSTSSLADYFWWIVAFSAMPLLLVLSAVLARYVILLK
	orf64a	10 20 30 40 50 60 MRRFLPIAAICAVLLYGLTAATGSTSSLADYFWWIVAFSAMPLLLVLSAVLARYVILLK
10	orf64.pep	70 80 90 100 110 120 DRRDGVFGSXXAKXPXXMETLVAXLPGVFLFGFPAQFINGTINSWFGNDTHEALERSLN
	orf64a	70 80 90 100 110 DRRDGVFGSQAIR-LSGMETLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLN
15	orf64.pep	130 140 150 160 170 180 LSKSALNLAADNALGNAV PVQIDLIGAASLP GDMGRVLEHYAGSGFAQLALYNXASGKIE
	orf64a	120 130 140 150 160 170 LSKSALNLAADNALGNAIPVQIDXIGAASLPXDMGRVLEHYAGSGFAQLALYNAASGKIE
20	orf64.pep	190 200 210 220 230 240 KSINPHKLDQPFPGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHXGRDYALFFRQP
	orf64a	180 190 200 210 220 230 KSINPHKLDQPFPGKARWEKIQAGSVRDXESIGGVLYAXGWLSAXTHNGRDYALFFRQP
30	orf64.pep	250 260 270 280 290 300 VPKGVAEDAVLIEKARAKYAELSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFV
	orf64a	240 250 260 270 280 290 VPKGVAEDAVLIEKARAXXXLSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFV
35	orf64.pep	310 320 330 340 350 360 EPVLSLAEGAKAVAQGD FSQTRPVLNRNDEFGRLTXLFN HMTQLSIAKDADERNRRREEA
	orf64a	300 310 320 330 340 350 EPVLSLAEGAKAVAQGD FSQTRPVLNRNDEFGRLTKLFN HMTQLSIAKEADERNRRREEA
40	orf64.pep	370 380 390 ARHYLECVLEGLTTGVVVFDEQGCLKTFNKAAGT
	orf64a	360 370 380 390 400 410 ARHYLECVLEGLTTGVVVFDEQGCLKTFNKA AEQILGMPLTPLWGSSRRHGWSAQQSL
45	orf64a	420 430 440 450 460 470 LAEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNXGVMVIDDITVLIHAQ

The complete length ORF64a nucleotide sequence <SEQ ID 253> is:

50	1	ATGCGCCGTT	TTCTACCGAT	CGCAGCCATA	TGCGCCGTCG	TCCTGTTGTA
	51	CGGACTGACG	GCGGCAACCG	GCAGCACCAG	TTCGCTGGCG	GATTATTTCT
	101	GGTGGATTGT	TGCGTTTCAGC	GCAATGCTGC	TGCTGGTGT	GTCCGCCGTT
	151	TTGGCAGGTT	ATGTCATATT	GCTGTTGAAA	GACAGGCGCG	ACGGCGTATT
55	201	CGGTTTCGAG	ATTGCCAAAC	GCCTTTCGGG	GATGTTTACG	CTGGTTGCCG
	251	TACTGCCCGG	CGTGTTCCTG	TTCGCGGTTT	CCGCACAGTT	TATCAACGGC
	301	ACGATTAATT	CGTGGTTCGG	CAACGATACC	CACGAGGCGC	TTGAACGCAG
	351	CCTCAATTTG	AGCAAGTCCG	CATTGAATCT	GGCGGCAGAC	AACGCCCTTG
60	401	GCAACGCCAT	CCCCGTGCAG	ATAGACNTCA	TGGCGCGCGC	TTCCCTGCCC
	451	NGGGATATGG	GCAGGGTGCT	GGAACATTAC	GCCGGCAGCG	GTTTTGCCCA
	501	GCTTGCCCTG	TACAATGCCG	CAAGCGGCAA	AATCGAAAAA	AGCATCAACC
	551	CGCACAAAGCT	CGATCAGCCG	TTTCCAGGTA	AGGCGCGTTG	GGAAAAAATC
65	601	CAACAGGCGG	GTCGGTCTAG	GGATNNGGAA	AGCATAGGCG	GCATATTGTA
	651	CGCGCANGGC	TGGCTGTCTG	CAGNNACGCA	CAACGGGCGC	GATTACGCCT
	701	TGTTTTTCCG	TCAGCCGGTT	CCCAAAGGCG	TGGCAGAGGA	TGCCGTCTTA
	751	ATCGAAAAGG	CAAGGGCGNA	ANANNNTNAG	TTGAGTTACA	GCAAAAAAGG
	801	TTTGACAGAC	TTTTTCTCTG	CAACCCTGCT	GATTGCCTCN	CTGCTGTCGA
	851	TTTTTCTTGC	ACTGGTCATG	GCACTGTATT	TCGCCCGCGG	TTTCGTTCGA

5  
 10  
 15  
 20  
 25

```

901 CCCGTCCTAT CGCTTGCCGA GGGGGCGAAG GCGGTGGCGC AAGGCGATT
951 CAGCCAGACG CGCCCCGTGT TGCGCAACGA CGAGTTCGGA CGCTTGACCA
1001 AGTTGTTCAA CCACATGACC GAGCAGCTTT CCATCGCCAA AGAAGCAGAC
1051 GAGCGCAACC GCCGGCGCGA GGAAGCCGCC AGACATTATC TCGAATGCGT
1101 GTTGGAGGGG CTGACCACGG GCGTGGTGGT GTTTGACGAA CAAGGCTGTC
1151 TGA AACCTT CAACAAAGCG GCGGAACAGA TTTTGGGGAT GCCGCTTACC
1201 CCCCTGTGGG GCAGCAGCCG GCACGGTTGG CACGGCGTTT CGGCGCAGCA
1251 GTCCCTGCTT GCCGAAGTGT TTGCCGCCAT CGGCGCGGCG GCAGGTACGG
1301 ACAAACCGGT CCATGTGAAA TATGCCGCGC CGGACGATGC CAAAATCCTG
1351 CTGGGCAAGG CAACCGTCTT GCCCGAAGAC AACNGCAACG GCGTGGTAAT
1401 GGTGATTGAC GACATCACCG TTTTGATACA CGCGCAAAA GAAGCCGCGT
1451 GGGGCGAAGT GGCAAAACGG CTGGCACACG AAATCCGCAA TCCGCTCACG
1501 CCCATCCAGC TTTCTGCCGA ACGGCTGGCG TGGAAATTGG GCGGGAAGCT
1551 GGACGAGCAN GACGCGCAAA TCCTGACACG TTCGACCGAC ACCATCATCA
1601 AACAAAGTGC GGCATTAAAA GAAATGGTCG AGGCATTCCG CAATTACNCG
1651 CGTTCCCTTT CGNCTCAATT GGAAATCAG GATTTGAACG CCTTAATCGG
1701 CGATGTGTTG GCATTGTACG AAGCTGGTCC GTGCCGTTT GCGGCGGAAC
1751 TTGCCGCGCA ACCGCTGATG ATGGCGGCGG ATACGACCGC CATGCGCAG
1801 GTGCTGCACA ATATTTTCAA AAATGCCGCC GAAGCGGCGG AAGAAGCCGA
1851 TGTGCCCGAA GTCAGGTAA AATCGGAAGC GGGGCAGGAC GGACGGATTG
1901 TCCTGACAGT TTGCGACAAC GGCAAGGGGT TCGGCAGGGA AATGCTGCAC
1951 AATGCCTTCG AGCCGTATGT AACGGACAAA CCGCTGGAA CGGGATTGNG
2001 ACTGCCGTG GTGAAAAAAA TCATTGAAGA ACACGGCGGC CNCATCAGCC
2051 TGAGCAATCA GGATGCGGGC GCGCGTNTG TCAGAATCAT CTTGCCAAAA
2101 ACGGTAGAAA CTTATGCGTA G
  
```

This encodes a protein having amino acid sequence <SEQ ID 254>:

30  
 35  
 40

```

1 MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV
51 LARYVILLK DRRDGVFGSQ IAKRLSGMFT LVAVLPGVFL FGVSAQFING
101 TINSWFGNDT HEALERSLNL SKSALNLAAD NALGNAIPVQ IDXIGAASLP
151 XDMGRVLEHY AGSGFAQLAL YNAASGKIEK SINPHKLDQP FPGKARWEKI
201 QQAGSVRDXE SIGGVLYAXG WLSAXTHNGR DYALFFRQPV PKGVAEDAVL
251 IEKARAXXXX LSYSKKGLOT FFLATLLIAS LLSIFLALVM ALYFARRFVE
301 PVLSLAEGAK AVAQGDFSQT RPYLRNDEFG RLTKLFNHMT EQLSIAKEAD
351 ERNRRREEAA RHYLECVLEG LTTGVVVFDE QGCLKTFNKA AEQILGMPLT
401 PLWGSSRHGW HGVSAAQSLL AEFVFAIGAA AGTDKPVHVK YAAPDDAKIL
451 LGKATVLPED NXNGVMVID DITVLIHAQK EAAWGEVAKR LAHEIRNPLT
501 PIQLSAERLA WKLGGKLDX DAQILTRSTD TIIKQVAALK EMVEAFRNYX
551 RSPSXQLENQ DLNALIGDVL ALYEAGPCRF AAELAGEPLM MAADTTAMRQ
601 VLHNIFFKNA EAAEADVPE VRVKSEAGQD GRIVLTVCDN KGKFGREMLH
651 NAFEPYVTDK PAGTGLXLPV VKKIIIEHGG XISLSNQDAG GAXVRIILPK
701 TVETYA*
  
```

ORF64a and ORF64-1 show 96.6% identity in 706 aa overlap:

45  
 50  
 55  
 60  
 65

```

              10      20      30      40      50      60
orf64a.pep  MRRFLPIAAICAVVLLYGLTAATGSTSSLDYFWWIVAFSAML VLSAVLARYVILLK
              |||
orf64-1      MRRFLPIAAICAVVLLYGLTAATGSTSSLDYFWWIVAFSAML VLSAVLARYVILLK
              |||
              10      20      30      40      50      60

              70      80      90      100     110     120
orf64a.pep  DRRDGVFGSQIAKRLSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLNL
              |||
orf64-1      DRRDGVFGSQIAKRLSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLNL
              |||
              70      80      90      100     110     120

              130     140     150     160     170     180
orf64a.pep  SKSALNLAADNALGNAIPVQIDXIGAASLPXDMGRVLEHYAGSGFAQLALYNAASGKIEK
              |||
orf64-1      SKSALNLAADNALGNAIPVQIDLIGAASLPXDMGRVLEHYAGSGFAQLALYNAASGKIEK
              |||
              130     140     150     160     170     180

              190     200     210     220     230     240
orf64a.pep  SINPHKLDQFPFGKARWEKIQAGSVRDLESIGGVLYAXGWLSAXTHNGRDYALFFRQPV
              |||
orf64-1      SINPHKLDQFPFGKARWEKIQAGSVRDLESIGGVLYAXGWLSAXTHNGRDYALFFRQPV
              |||
              190     200     210     220     230     240

              250     260     270     280     290     300
  
```

5	orf64a.pep	PKGVAEDAVLIEKARAXXXLSYSKGLQTFLLATLLIASLLSIFLALVMALYFARRFVE
	orf64-1	PKGVAEDAVLIEKARAKYAELSYSKGLQTFLLATLLIASLLSIFLALVMALYFARRFVE
10	orf64a.pep	PVLSLAEGAKAVAQGDfsQTRPVLNRNDEFGRLTKLFNHMTQLSIAKEADERNRRREEAA
	orf64-1	PVLSLAEGAKAVAQGDfsQTRPVLNRNDEFGRLTKLFNHMTQLSIAKEADERNRRREEAA
15	orf64a.pep	RHYLECVLEGLTTGVVVfDEQGCLKTfNKAAEQILGMPLTfPLWGSSRHGWHGVSAQQSLL
	orf64-1	RHYLECVLEGLTTGVVVfDEQGCLKTfNKAAEQILGMPLTfPLWGSSRHGWHGVSAQQSLL
20	orf64a.pep	AEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNNGVVMVIDDITVLIHAQK
	orf64-1	AEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNNGVVMVIDDITVLIHAQK
25	orf64a.pep	EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLDEXDAQILTRSTDITIKQVAALK
	orf64-1	EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLDEXDAQILTRSTDITIKQVAALK
30	orf64a.pep	EMVEAFRNYKRSPSXQLENQDLNALIGDVLALYEAGPCRFAELAGEPLMMAADTTAMRQ
	orf64-1	EMVEAFRNYKRSPSLKLENQDLNALIGDVLALYEAGPCRFAELAGEPLTVAADTTAMRQ
35	orf64a.pep	VLHNI fKNAAEAAEEADVPEVRVKSEAGQDGRIVLTVCNKGKFGREMLHNAFEPYVTDK
	orf64-1	VLHNI fKNAAEAAEEADVPEVRVKSETGQDGRIVLTVCNKGKFGREMLHNAFEPYVTDK
40	orf64a.pep	PAGTGLXLPVVKKIIIEHGGXISLSNQDAGGAXVRIILPKTVETYAX
	orf64-1	PAGTGLGLPVVKKIIIEHGGRIISLSNQDAGGACVRIILPKTVKTYAX

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF64 shows 86.6% identity over a 387aa overlap with a predicted ORF (ORF64.ng) from *N.*

### 50 *gonorrhoeae*:

55	orf64.pep	MRRFLPIAAICAXLXXGLTAATGSTSSLADYFWWIVAFSAMPLLLVLSAVLARYVILLK	60
	orf64ng	MRRFLPIAAICAVVLLYGLTAATGSTSSLADYFWWIVSFSAMPLLLVLSAVLARYVILLK	60
60	orf64.pep	DRRDGVFGSXXAKXPXXMFTLVAXLPGVFLFGFPAQFINGTINSWFGNDTHEALERSLN	120
	orf64ng	DRRNGVFGSQIAKR-LSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLN	119
65	orf64.pep	LSKSALNLAADNALGNAPVPQIDLIGAASLPGDMGRVLEHYAGSGFAQLALYNXASGKIE	180
	orf64ng	LSKSALDLAADNAVSNAPVPQIDLIGTASLSGNMGVSLEHYAGSGFAQLALYNAASGKIE	179
65	orf64.pep	KSINPHKLDQFPFGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHXGRDYALFFRQP	240
	orf64ng	KSINPHQFDQPLPDKEHWEIQQTGSVRSLESIGGVLYAQGWLSAGTHNGRDYALFFRQP	239



```

15      1  MRRFLPIAAI  CAVVLLYGLT  AATGSTSSLA  DYFWWIVSFS  AMLLLVLVSAV
      51  LARYVILLLK  DRRNGVFGSQ  IAKRLSGMFT  LVAVLPGFLF  FGISAQFING
      101  TINSWFGNdT  HEALERSLNL  SKSALDLAAD  NAVSNAPVPO  LDIGTASLS
      151  GNMGSVLEHY  AGSGFAQLAL  YNAASGKIEK  SINPHQFDQP  IDPKEHWEQI
      201  QQTGSVRSLE  SIGGVLYAQG  WLSAGTHNGR  DYALFFRQPI  PENVAQDAVL
      251  IEKARAKYAE  LSYSKKGLQT  FFLVTLIIAS  LLSIFLALVM  ALYFARRFEV
20      301  PILSLAEGAK  AVAQGDVSQT  RPVLNRNVEFG  RLTKLFNHMT  EQLSIAKEAD
      351  ERNRRREEAA  RHYLECVDLG  LTTGVVVSYF  LSCCRTAVFS  TCHSSPLSYF+

```

	1	ATGCGCCGCT	TCCTACCGAT	CGCAGCCATA	TGCGCCGTCG	TCCTGCTGTA
25	51	CGGATTGACG	GCGGCGACCG	GCAGCACCAG	TTCTGCTGGCG	GATTATTTCT
	101	GGTGGATAGT	CTCGTTCAGC	GCAATGCTGC	TGCTGGTGTT	GTCCGCGGTT
	151	TTGGCACGTT	ATTGCATATT	GCTGTTGAA	GACAGCGGCA	ACGGCGTGTT
	201	CGGTTCCGAG	ATTGCCAAAC	GCCTTTCGG	GATGTTCAAG	CTGGTCGCCG
	251	TACTGCCCGG	CTTGTTCTCT	TTCGGCATT	CCGCGCAGTT	TATCAACGGC
30	301	ACGATTAATT	CGTGGTTCGG	CAACGACACC	CACGAAGCCC	TGCAACGCGA
	351	CCTTAATTTG	AGCAAAGTCC	CACCTGGATT	GGCGGCAGCC	AATGCCGTCA
	401	GCAACGCCGT	TCCCGTACAG	ATAGACCTCA	TCCGGACCGC	CTCCTGTCTG
	451	GGCAATATGG	GCAGTGTGCT	GGAACACTAC	GCCGGCAGCG	GTTTTGCCCA
	501	GCTTGCCCTG	TACAATGCCG	CAAGCGGGAA	AATCGAAAAA	AGCATCAATC
35	551	CGCACCAATT	CGACCAGCCG	CTTCCCGACA	AAGAACAATT	GGAACAGATT
	601	CAGCAGACCG	GTTCCGTTCT	GAGTTTGGA	AGCATAGGCG	CGGTATTGTA
	651	CGCGCAGGGA	TGGTTGTCTG	CAGGTACGCA	CAACGGGCGC	GATTACGCGC
	701	TGTTCTTCCG	CCAGCCGATT	CCCGAAAAAT	TGGCAGAGGA	TGCCGTTCTG
	751	ATTGAAAAGG	CGCGGGCGAA	ATATGCCGAA	TTGAGTTACA	GCAAAAAGG
40	801	TTTGAGAAC	TTTTTCTTGG	TAACCTGTCT	GATTGCTCTG	TCGCTGTCTG
	851	TTTTTCTTGC	GCTGGTAATG	GCACTGTATT	TTGCCCGCGC	TTTCGTCTGA
	901	CCATTCTGT	CGCTTGCCGA	GGCGCGAAAG	GCGGTGGCGC	AGGGTGATTT
	951	CAGCCAGACG	CGCCCGGTAT	TGCGCAACGA	CGAGTTCGGA	CGTTTGACCA
	1001	AGCTGTTCAA	CCATATGACC	GAGCAGCTTT	CCATGCGCAA	AGAAGCAGAC
45	1051	GAACGCCAACC	GCCGGCGCGA	GGAAGCCGCC	CGTCACTACC	TCGAGTGCCT
	1101	GTTGGATGGG	TTGACTACCG	GTGTGGTGGT	GTTTGACGAA	AAAGGCCGTT
	1151	TGAAAACCTT	CAACAAGGCG	GCGGAACAGA	TTTTGGGGAT	GCCGCTCGCC
	1201	CCCTGTGGG	GCAGCAGCCG	GCACGGTTGG	CACGGCGTTT	CGGCGCAGCA
	1251	GTCCTGCTT	GCCGAAGTGT	TtgcgcgcAT	CGGTGGCGGC	GCAGTACCGG
50	1301	ACAAACCGGT	CCAGGTGGAA	TATGCCGCGC	CGGACGATGC	CAAAATCCTG
	1351	CTGGGCAAGG	CGACGGTATT	GCCCGAAGAC	AACGGCAACG	GCGTGGTGAT
	1401	GGTGATTGAC	GACATCACCG	TGCTGATACG	CGCGCAAAAA	GAAGCCGCGT
	1451	GGGGTGAAGT	GGCGAAGCGG	CTGGCACACG	AAATCCGCAA	TCCGCTCACG
	1501	CCCATCCAGC	TTTCCGCCGA	ACGGCTGGCG	TGGAATTGG	GCGGGAAGCT
	1551	GGACGATCAG	GACGCGCAAA	TCCTGACGCG	TtcgACCGAC	ACCATCATCA
55	1601	AACAGgtggc	gCGCTTAAAA	GAATGGTCTG	AGGCATTCCG	CAATTACGGC
	1651	CGCGCCCTTT	GCGTCAAACT	GGAAAATCAG	GATTGTAAAG	CCTTAATCGG
	1701	CGATGTTTTG	GCCCTGTACG	AAGCCGCCCC	GTGCCGGTTT	GAGGCGGAAC
	1751	TTGCCGGCGA	ACCGCTGATG	ATGGCGGCGG	ATACGACCGC	CATGCGGCAG
60	1801	GTGCTGCACA	ATATTTTCAA	AAATGCCGCC	GAAGCGGCGG	AAGAAGCCGA
	1851	TATGCCCGAA	GTCAAGGTAA	AATCGGAAAC	GGGCGAGGAC	GGACGAGATT
	1901	TCTGACGGT	TGCGACAAC	GGCAAGGGAT	TCCGCAAGGA	AATGCTGCAC
	1951	AATGCTTTCT	AGCCGTATGT	GACGGATAAG	CCGGCGGGAA	CGGGACTGGG
	2001	TCTGCCTGTA	GTGAAAAAAA	TCATTGGAGA	ACACGGCGGC	CGCATCAGCC
	2051	TGAGCAATCA	GGATCGGGGT	GGGGCGTGTG	TCGAATCAT	CTTGCCAAAA
65	2101	ACGGTGAATA	CTTATGCGTA	G		

5

10

15

20

2

3

3.

•

10

15

Furthermore, ORF64ng-1 shows significant homology to a protein from *A. caulinodans*:

sp|Q04850|NTRY\_AZOCA NITROGEN REGULATION PROTEIN NTRY >gi|77479|pir||S18624 ntrY  
 protein - Azorhizobium caulinodans >gi|38737 (X63841) NtrY gene product  
 [Azorhizobium caulinodans] Length = 771  
 Score = 218 bits (550), Expect = 7e-56  
 Identities = 195/720 (27%), Positives = 320/720 (44%), Gaps = 58/720 (8%)

Query: 7 IAAICAVVLLYLTAATGSTSSLDYFWWIXXXXXXXXXXXXXXXXXXRYVILLKDRNGV 66  
I+A+ ++L GLT + + + R + + K R G  
Sbjct: 35 ISALATFLILMGLTPVVPVTHQVVIS----VLLVNAAAVLILSAMVGREIWRIAKRARGR 90

Query: 67 FGSQIAKRLSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLNLSKSALD 126  
 +++ R+ G+F +V+V+P + + +++ ++ ++ WF TE + S+++++ +  
 Sbjct: 91 AAARLHIRIVGLFAVVSVPVAILVAVVASLTLDRLGDRWFMSMRTOEIVASSVSVAQTYVR 150

Query: 127 LAADNAVSNVAVPQIDLIGTASLSGNMGSVLEHYAG--SGFAQLALYNAASGKIEKSINP 184  
 A N + + + DL S+ Y G S F Q+ AA + ++  
 Sbjct: 151 EHAIINIRGDILAMSADLTRLKSV-----YEGDRSRFNQILTAQAALRNLPGLMI 200

Query: 185 HQFDQPLPDKEHWEQIQQTGSVRSLESIGGVLYAQGWLSAGTHNGRDYA----- 233  
 + D + + + + I + V + +IG Q + N DY  
 Sbjct: 201 RR-DLSVVERAN-VNIGREFIVPANLAIGDATPDOPVIYLP--NDADYVAAVVPLKDYDD 256

Query: 234 --LFRQPIPENVAQDAVLIEKARAKYAELSYSKKLQTFFLVTXXXXXXXXXXXXVMA 291  
L+ + I V ++ A Y L + G+Q F + +  
Sbjct: 257 LYLYVARLIDPRVIGYLKTTQETLADYRSLEERFVGVOVAFALMYAVITLIVLLSAVWL 316

Query: 292 LYFARRFVEPILSLAEGAKAVAQGDFSQTRPVLRLND-EFGRLTKLFNHMTEQLSIXXXX 350  
L F++ V PI L A VA+G+ P+ R + + L + FN MT +L  
Sbjct: 317 LNFSKWLVAPIRRLMSAADHVAEGNLDVRVPIYRAEGDLASLAETENKMTHELRSOREAI 376

Query: 351 XXXXXXXXXXXXHYLECVLDGLTTGVVVFDEKGRCLKTFNKAAEQILGMLAPLWGSRRHW 410  
+ E VL G+ GV+ D + R+ N++AE++LG L+ + RH  
Sbjct: 377 LTARDQIDSRRTFEAVLSGVGAGVIGLDSOERITILNRSAPRLG--LSEVEALRRHLA 434

Query: 411 HGVSAAQSLLEAEVFXXXXXXXTDKPVQVEYAAPDDAKILLGKATVLPEDNG---NGVVM 467  
V LL E + VQ D + + V E + +G V+

Sbjct: 435 EVVPETAGLLEEA-----EHARORSVOGNITLTRDGRERVFVVRVTTEOSPEAHGWWV 488

Query: 468 VIDDITVLIRAQKEAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGKLLDDQDAQILTR 527  
+DDIT LI AQ+ +AW +VA+R+AHEI+NPLTPIQLSAERL K G + QD +I +  
Sbjct: 489 TLDDITELISAORTSAWADVARRIAHEIKNPLTPIQLSAERLKRKEGRHV-TDREIFEDQ 547

Query: 528 STDIIKQVAALKEMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFEAEELAGE 587  
 TDII+QV + MV+ F ++AR P +++QD++ +I + L G +  
 Sbjct: 548 CTDIIROVGDIGRMVDFSSFARMKPVPVDSOMSEIIROTVELMRVGHPEVVEVDFSEVP 607

Query: 588 PLMAA-DTTAMRQVLHNIFKNXXXXXXXXDMPEVRVK-----SETGQDGRIVLTVCD 639  
P M A . D + Q L N I K N P + V R + + G + D + V + + D  
Sbjct: 608 PAMPARFDRRLVSQALTNLKNAEAIEAVP-PDVRGQGRIRVSANRVGED--LVIDIID 664

Query: 640 NGKGFGKEMLHNAFEPYVTDKPAAGTGLGLPVVKIIGEHGGRISLSNQDAG-GACVRIIL 698  
 NG G +E + EPYVT + GTGLGL +V KI+ EHGG I L++ G GA +R+ L  
 Sbjct: 665 NGTGLPQESRNLLEPYVTTREKGTGLGLAIVGKIMEEHGGIELNDAPEGRGAWIRLTL 724

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### 10 Example 31

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 259>:

```

1 ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT
51 GCTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
101 CTTTCCAAAT TTTCGGCATC CACACCACTT GGGGCGCATT TTCCTTTCCC
151 TTCATCTTCC TTGCCACCGA CCTGACCGTC CGCATTTTCG GTTCTCACTT
201 GGCACGGCGG ATTATCTTTT GGGTGATGTT CCCC GCCCTT TTGCTTTTCT
251 ACGTCTTTTC CGTTTTGTTC CACAACGGCA GTTGGACAGG CTGGGGCGCG
301 CTGTCCGAAT TCAACACCTT TGTCGGACGC ATCGCCTTAG CCAGCTTTGC
351 CGCCTACGCG ATCGGACAAA TCCTTGATAT TTTTGTATT CACAAATTAC
401 GCCGTCTGAA AGCGTGGTGG ATTGCACCGA ACGCATCAAC CGTCATCGGG
451 CACGCGTTGG ATACG...

```

This corresponds to the amino acid sequence <SEQ ID 260; ORF66>:

```

1 MYAFTAAQQQ KALFRLVLFH ILIIAASNYL VQFPFQIFGI HTTWGAFSFP
51 FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA
101 LSEFNTFVGR IALASFAAYA IGQILDIFVF NKLRLRKAWW IAPNASTVIG
151 HALDT...

```

Further work revealed the complete nucleotide sequence <SEQ ID 261>:

```

1 ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT
51 GCTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
101 CTTTCCAAAT TTTCGGCATC CACACCACTT GGGGCGCATT TTCCTTTCCC
151 TTCATCTTCC TTGCCACCGA CCTGACCGTC CGCATTTTCG GTTCTCACTT
201 GGCACGGCGG ATTATCTTTT GGGTGATGTT CCCC GCCCTT TTGCTTTTCT
251 ACGTCTTTTC CGTTTTGTTC CACAACGGCA GTTGGACAGG CTGGGGCGCG
301 CTGTCCGAAT TCAACACCTT TGTCGGACGC ATCGCCTTAG CCAGCTTTGC
351 CGCCTACGCG ATCGGACAAA TCCTTGATAT TTTTGTATT CACAAATTAC
401 GCCGTCTGAA AGCGTGGTGG ATTGCACCGA CCGCATCAAC CGTCATCGGC
451 AACGCGTTGG ATACGCTGGT ATTTTTCGCC GTTGCCTTCT ACGCAAGCAG
501 CGATGGATT ATGGCGGCAA ACTGGCAGGG CATCGCTTTT GTCGATTACC
551 TGTTCAAAC TACCGTCTGC ACCCTCTTCT TCCTGCCCCG CTACGGCGTG
601 ATACTGAATC TGCTGACGAA AAAACTGACA ACCCTGCAAA CCAAACAGGC
651 GCAAGACCGC CCCGCGCCCT CGCTGCAAAA TCCGTAA

```

This corresponds to the amino acid sequence <SEQ ID 262; ORF66-1>:

```

1 MYAFTAAQQQ KALFRLVLFH ILIIAASNYL VQFPFQIFGI HTTWGAFSFP
51 FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA
101 LSEFNTFVGR IALASFAAYA IGQILDIFVF NKLRLRKAWW IAPTASTVIG
151 NALDTLVFFA VAFYASSDGF MAANWQGI AFVDYLFKLTVC TLFFLPAYGV
201 ILNLLTKKLT TLQTKQAQDR PAPSLQNP*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical protein o221 of *E. coli* (accession number P37619)

ORF66 and o221 protein show 67% aa identity in 155aa overlap:

orf66 1 MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSEFPFIFLATDLTV 60  
 M F+ Q+ KALF L LFH+L+I +SNYLQ P I G HTTWGAFSEFPFIFLATDLTV  
 o221 1 MNVFSQTQRYKALFWLSLFLHLLVITSSNYLVQLPVSIILGFHTTWGAFSEFPFIFLATDLTV 60

5 orf66 61 RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA 120  
 RIFG+ LARRIIF VM PALL+SYV S LF+ GSW G GAL+ FN FV RIA ASF AYA  
 o221 61 RIFGAPLARRIIFAVMIPALLISYVISSLFYMGSWQGFALAHFNLFVARIATASF MAYA 120

10 orf66 121 IGQILDIFVFNKLRLKAWWIAPNASTVIGHALDT 155  
 +GQILD+ VFN+LR+ + WW+AP AST+ G+ DT  
 o221 121 LGQILDVHVFNRLRQSRWWLAPTASTLFGNVSDT 155

### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF66 shows 96.1% identity over a 155aa overlap with an ORF (ORF66a) from strain A of *N.*

15 *meningitidis*:

		10	20	30	40	50	60
orf66.pep		MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSEFPFIFLATDLTV					
orf66a		MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQISGIHTTWGAFSEFPFIFLATDLTV					
20		10	20	30	40	50	60
		70	80	90	100	110	120
orf66.pep		RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA					
orf66a		RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA					
25		70	80	90	100	110	120
		130	140	150			
orf66.pep		IGQILDIFVFNKLRLKAWWIAPNASTVIGHALDT					
		:					
orf66a		LGQILDIFVFNKLRLKAWWVAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGIAF					
		130	140	150	160	170	180
30							
orf66a		VDYLEFKLTVCGLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX					
		190	200	210	220		
35							

The complete length ORF66a nucleotide sequence <SEQ ID 263> is:

	1	ATGTACGCAT	TTACCGCCGC	ACAGCAACAG	AAGGCACTCT	TCTGGCTGGT
	51	GCTTTTTCAT	ATCCTCATCA	TCGCCGCCAG	CAACTATCTG	GTGCAGTTCC
40	101	CCTTCCAAAT	TTCCGGGCATC	CACACCACTT	GGGGCGCGTT	TTCTTTCCC
	151	TTCATCTTCC	TCGCCACCGA	CCTGACCGTC	CGCATTTCG	GTTCCGACTT
	201	GGCAGGCGG	ATTATCTTTT	GGGTCATGTT	CCCCGCCCTT	TTGCTTTCCT
	251	ACGTCTTTC	CGTTTGTTC	CACAACGGCA	GTTGGACGGG	CTTGGGCGCG
	301	CTGTCCGAAT	TCAACACCTT	TGTCGGACGC	ATCGCGCTGG	CAAGTTTTC
45	351	CGCCTACGCG	CTCGGACAAA	TCCTTGATAT	TTTGTGTTC	AACAAATTAC
	401	GCCGTCTGAA	AGCGTGGTGG	GTTGCCCCGA	CTGCATCAAC	CGTCATCGGC
	451	AACGCCTTAG	ATACGTTGGT	ATTTTTCGCC	GTTGCCTTCT	ACGCAAGCAG
	501	CGATGGATT	ATGGCGGCAA	ACTGGCAGGG	CATCGCTTTT	GTCGATTACC
	551	TGTTCAAAC	CACCGTCTGC	GGTCTGTTT	TCCTGCCCGC	CTACGGCGTG
50	601	ATTCTGAATC	TGCTGACGAA	AAACTGACG	ACCCTGCAAA	CCAAACAGGC
	651	GCAAGACCGC	CCCGCGCCCT	CGCTGCAAAA	TCGTAA	

This encodes a protein having amino acid sequence <SEQ ID 264>:

	1	MYAFTAAQQQ	KALFWLVLFH	ILIIAASNYL	VQFPFQISGI	HTTWGAFSFP
	51	FIFLATDLTV	RIFGSHLARR	IIFWVMFPAL	LLSYVFSVLF	HNGSWTGLGA
55	101	LSEFNTFVGR	IALASFAAYA	LGQILDIFVF	NKLRLKAWW	VAPTASTVIG
	151	NALDTLVFFA	VAFYASSDGF	MAANWQGIAF	VDYLEFKLTV	GLFFLPAYGV
	201	ILNLLTKKLT	TLQTKQAQDR	PAPSLQNP*		

ORF66a and ORF66-1 show 97.8% identity in 228 aa overlap:

		10	20	30	40	50	60
orf66a.pep		MYAFTAAQQQKALFWLVLFHILIIAASNYLVQFPFQISGIHTTWGAFSFPFIFLATDLTV					
orf66-1		MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV					
60							

-190-

		10	20	30	40	50	60
		70	80	90	100	110	120
5	orf66a.pep	RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA					
	orf66-1	RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA					
		70	80	90	100	110	120
10	orf66a.pep	LGQILDIFVFNKLRLKAWWVAPTASTVIGNALDTLVFFAVAFYASSDGEMAANWQGI AF					
	orf66-1	IGQILDIFVFNKLRLKAWWIAPTASTVIGNALDTLVFFAVAFYASSDGEMAANWQGI AF					
		130	140	150	160	170	180
15	orf66a.pep	VDYLFKLTVCGLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX					
	orf66-1	VDYLFKLTVCGLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX					
		190	200	210	220	229	

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF66 shows 94.2% identity over a 155aa overlap with a predicted ORF (ORF66.ng) from *N. gonorrhoeae*:

25	orf66.pep	MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV	60
	orf66ng	MYALTAQQQKALFRLVLFHILIIAASNYLVQFPFRIFGIHTTWGAFSFPFIFLATDLTV	60
30	orf66.pep	RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA	120
	orf66ng	RIFGSHLARRIIFWVMFPALSLSYVFSVLFHNGSWTGLGAPSQFNTFVGRIALASFAAYA	120
	orf66.pep	IGQILDIFVFNKLRLKAWWIAPNASTVIGHALDT	155
	orf66ng	LGQILDIFVFDKLRLKAWWIAPASTVIGNALDTLVFFAVAFYASSDEFMAANWQGI AF	180

35 The complete length ORF66ng nucleotide sequence <SEQ ID 265> is:

```

1  ATGTACGCAT TGACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT
51  GCTTTTCCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
101 CCTTCCGAT TTTCGGCATC CACACCACTT GGGGCGCGTT TTCCTTTCCC
151 TTCATCTTCC TCGCCACCGA CCTGACCGTC CGCATTTTCG GTTCGCACTT
201 GCGCGGCGG ATTATCTTTT GGGTGATGTT CCCC GCCCTT ttgCTTTcat
251 aCGTCTTTTC CGTTTTGTTC CACAACGGCA GTTGGACGGG CTTGGGCGCG
301 ctgTCCCAAT TCAACACCTT TGTCGGACGC ATCGCGCTGG CAAGTTTTCG
351 CGCCTACGCG CTCGGACAAA TCCTTGATAT TTTCTGATTG GACAAATTAC
401 GCCGTCTGAA AGCGTGGTGG ATTGCCCCGG CCGCATCAAC CGTCATCGGC
451 AATGCACTGG ACACGTTAGT ATTTTGTGCC GTTGCCTTTT ACGCAAGCAG
501 CGATGAATTT ATGGCGGCAA ACTGGCAGGG CATCGCTTTT GTCGATTACC
551 TGTTCAAATC TACCGTCTGC ACCCTCTTCT TCCTGCCCGC CTACGGCGTG
601 ATACTGAATC TGCTGACGAA AAAACTGACG GCCCTGCAAA CCAACAGGC
651 GCAAGACCGC CCCGTGCCCT CGCTGCAAAA TCCGTAA

```

50 This encodes a protein having amino acid sequence <SEQ ID 266>:

```

1  MYALTAQQQ KALFRLVLFH ILIIAASNYL VQFPFRIFGI HTTWGAFSFP
51  FIFLATDLTV RIFGSHLARR IIFWVMFPAL SLSYVFSVLF HNGSWTGLGA
101 PSQFNTFVGR IALASFAAYA LGQILDIFVF DKLRLKAWW IAPAASTVIG
151 NALDTLVFFA VAFYASSDEF MAANWQGI AF VDYLEFKLTVC TLFFLPAYGV
201 ILNLLTKKLT ALQTKQAQDR PVPSLQNP*

```

An alternative annotated sequence is:

```

1  MYALTAQQQ KALFRLVLFH ILIIAASNYL VQFPFRIFGI HTTWGAFSFP
51  FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA
101 LSQFNTFVGR IALASFAAYA LGQILDIFVF DKLRLKAWW IAPAASTVIG
151 NALDTLVFFA VAFYASSDEF MAANWQGI AF VDYLEFKLTVC TLFFLPAYGV
201 ILNLLTKKLT ALQTKQAQDR PVPSLQNP*

```

ORF66ng and ORF66-1 show 96.1% identity in 228 aa overlap:

```

5  orf66-1.pep  MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFIQIFGIHTTWGAFSFPFIPLATDLTV 60
   orf66ng     MYALTAAQQQKALFRLVLFHILIIAASNYLVQFPFRIQIFGIHTTWGAFSFPFIPLATDLTV 60

   orf66-1.pep  RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA 120
   orf66ng     RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSQFNTFVGRIALASFAAYA 120

10  orf66-1.pep  IGQILDIFVFNKLRRLLKAWWIAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGI AF 180
   orf66ng     LGQILDIFVFDKLRRLLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGI AF 180

15  orf66-1.pep  VDYLFKLTVCTLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX 229
   orf66ng     VDYLFKLTVCTLFFLPAYGVILNLLTKKLTALQTKQAQDRVPVPSLQNPX 229

```

Furthermore, ORF66ng shows significant homology with an *E.coli* ORF:

```

20  sp|P37619|YHHQ_ECOLI HYPOTHETICAL 25.3 KD PROTEIN IN FTSY-NIKA INTERGENIC
   REGION (O221)
   >gi|1073495|pir||S47690 hypothetical protein o221 - Escherichia coli >gi|466607
   (U00039) No definition line found [Escherichia coli] >gi|1789882 (AE000423)
   hypothetical 25.3 kD protein in ftsY-nikA intergenic region [Escherichia coli]
   Length = 221
   Score = 273 bits (692), Expect = 5e-73
25  Identities = 132/203 (65%), Positives = 155/203 (76%)

   Query: 1 MYALTAAQQQKALFRLVLFHILIIAASNYLVQFPFRIQIFGIHTTWGAFSFPFIPLATDLTV 60
   M + Q+ KALF L LFH+L+I +SNYLVO P I G HTTWGAFSFPFIPLATDLTV
   Sbjet: 1 MNVFSQTQRYKALFWLSLFHLLVITSSNYLVQLPVSILGFHTTWGAFSFPFIPLATDLTV 60

30  Query: 61 RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSQFNTFVGRIALASFAAYA 120
   RIFG+ LARRIIF VM PALL+SYV S LF+ GSW G GAL+ FN FV RIA ASF AYA
   Sbjet: 61 RIFGAPLARRIIFAVMIPALLISYVISSLFYMGSWQGFALAHFNLFVARIATASF MAYA 120

35  Query: 121 LGQILDIFVFDKLRRLLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGI AF 180
   LGQILD+ VF++LR+ + WW+AP AST+ GN DTL FF +AF+ S D FMA +W IA
   Sbjet: 121 LGQILDVHVFNRLQSRRLWLAAPTASTLFGNVSDTLAFFFFIAFWRSPDAFMAEHWMEIAL 180

40  Query: 181 VDYLFKLTVCTLFFLPAYGVILN 203
   VDY FK+ + +FFLP YGV+LN
   Sbjet: 181 VDYCFKVLISIVFFLPYMGVLLN 203

```

Based on this analysis, including the homology with the *E.coli* protein and the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 32

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 267>:

```

50  1 ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
   51 AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAyGCA GTmwrAATAT
   101 CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT
   151 GTACCTAAAA ATAGTAAAC TTATTCATCT GATTAATAA AAACGGTAGA
   201 TTTAACACAC AyyCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
   251 CCGCCAGCGT ATCCCGCGCC GCGGTATTGG CGGGGGTCGG CAACTTGCC
55  301 CGCTTAGGCG CGAAATTCAG CACAAGGGCG GTtCCCTATG TCGGAACAGC
   351 CcTTTTAGCC CACGACGTAT ACGAAAcTT CAAAGAAGAC ATACAGGCAC
   401 GAGGCTACCA ATACGACCCC GAAACCGACA AATTGTAAA AGGCTACGAA
   451 TATAGTAATT GCCTTGGTA CGAAGACAAA AGACGTATTA ATAGAACCTA

```

501 TGGCTGCTAC GCGTTGAT..

This corresponds to the amino acid sequence <SEQ ID 268; ORF72>:

5  
1 MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKEHKE  
51 VPKNSKTYSS DLIKTVDLTH XPTGAKARIN AKITASVSRA GVLAVGKLA  
101 RLGAKEFSTRA VPYVGTALLA HDVYETFKED IQARGYQYDP ETDKFKVGYE  
151 YSNCLWYEDK RRINRTYGCY GVD..

Further work revealed the complete nucleotide sequence <SEQ ID 269>:

10  
1 ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC  
51 AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAATGCA GTAAAAATAT  
101 CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT  
151 GTACCTAAAA ATAGTAAAAC TTATTCATCT GATTTAATAA AAACGGTAGA  
201 TTTAACACAC ATCCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA  
251 CCGCCAGCGT ATCCCGCGCC GCGGTATTGG CGGGGGTCGG CAAACTTGCC  
301 CGCTTAGGCG CGAAATTCAG CACAAGGGCG GTTCCCTATG TCGGAACAGC  
15  
351 CCTTTTAGCC CACGACGTAT ACGAACTTT CAAAGAAGAC ATACAGGCAC  
401 GAGGCTACCA ATACACCCC GAAACCGACA AATTGCAAA GGTCTCAGGC  
451 TAA

This corresponds to the amino acid sequence <SEQ ID 270; ORF72-1>:

20  
1 MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKEHKE  
51 VPKNSKTYSS DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGKLA  
101 RLGAKEFSTRA VPYVGTALLA HDVYETFKED IQARGYQYDP ETDKFAKVS  
151 \*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF72 shows 98.0% identity over a 147aa overlap with an ORF (ORF72a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orf72.pep		MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKEHKE VPKNSKTYSS					
30 orf72a		MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKEHKE VPKNSKTYSS					
		10	20	30	40	50	60
		70	80	90	100	110	120
orf72.pep		DLIKTVDLTH XPTGAKARIN AKITASVSRA GVLAVGKLARLGAKEFSTRA VPYVGTALLA					
35 orf72a		DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGKLARLGAKEFSTRA VPYVGTALLA					
		70	80	90	100	110	120
		130	140	150	160	170	
40 orf72.pep		HDVYETFKEDIQARGYQYDP ETDKFKVGYE YSNCLWYEDK RRINRTYGCY GVD					
orf72a		HDVYETFKEDIQARGYQYDP ETDKFAKVS GX					
		130	140	150			

The complete length ORF72a nucleotide sequence <SEQ ID 271> is:

45  
1 ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC  
51 AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAATGCA GTAAAAATAT  
101 CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT  
151 GTACCTAAAA ATAGTAAAAC TTATTCATCT GATTTAATAA AAACGGTAGA  
201 TTTAACACAC ATCCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA  
50  
251 CCGCCAGCGT ATCCCGCGCC GCGGTATTGG CGGGGGTCGG CAAACTTGCC  
301 CGCTTAGGCG CGAAATTCAG CACAAGGGCG GTTCCCTATG TCGGAACAGC  
351 CCTTTTAGCC CACGACGTAT ACGAACTTT CAAAGAAGAC ATACAGGCAC  
401 GAGGCTACCA ATACACCCC GAAACCGACA AATTGCAAA GGTCTCAGGC  
451 TAA

55 This encodes a protein having amino acid sequence <SEQ ID 272>:



-193-

```

1  MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF
51  VPKNSKTYSS DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGVKLA
101 RLGAKFSTRA VPYVGTALLA HDVYETFKED IQARGYQYDP ETDKFAKVS
151 *

```

5 ORF72a and ORF72-1 show 100.0% identity in 150 aa overlap:

```

10 orf72a.pep      10      20      30      40      50      60
    MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF VPKNSKTYSS
    |||||
orf72-1      10      20      30      40      50      60
    MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF VPKNSKTYSS

15 orf72a.pep      70      80      90     100     110     120
    DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGVK LARLGAKF STRAVPYV GTALLA
    |||||
orf72-1      70      80      90     100     110     120
    DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGVK LARLGAKF STRAVPYV GTALLA

20 orf72a.pep      130     140     150
    HDVYETFKED IQARGYQYDP ETDKFAKVS GX
    |||||
orf72-1      130     140     150
    HDVYETFKED IQARGYQYDP ETDKFAKVS GX

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

25 ORF72 shows 89% identity over a 173aa overlap with a predicted ORF (ORF72.ng) from *N. gonorrhoeae*:

```

30 orf72.pep      MVIKYTNLNF AKLSIIAILM MYSFEANANA VXiSETVSVD TGQGAKIHKF VPKNSKTYSS 60
    || :|||
orf72ng      MVTKHTNLNF AKLSIIAILM MYSFEANANA VKISETLSVD TGQGAKVHKF VPKSSNIYSS 60

orf72.pep      DLIKTVDLTH XPTGAKARIN AKITASVSRA GVLAVGVK LARLGAKF STRAVPYV GTALLA 120
    || :|||
orf72ng      DLTkAVDLTH IPTGAKARIN AKITASVSRA GVLsgvGKLVRQgAKFgTRAVPYV GTALLA 120

35 orf72.pep      HDVYETFKED IQARGYQYDP ETDKfVKGyEYsNCLWYEDKRrINRTYGCYGVd 173
    |||||
orf72ng      HDVYETFKED IQAGCRyDP ETDKfVKGyEYANCLWYEDERRINRTYGCYGVdSSIMRLM 180

```

An ORF72ng nucleotide sequence <SEQ ID 273> was predicted to encode a protein having amino acid sequence <SEQ ID 274>:

```

40 1  MVTKHTNLNF AKLSIIAILM MYSFEANANA VKISETLSVD TGQGAKVHKF
    51  VPKSSNIYSS DLTkAVDLTH IPTGAKARIN AKITASVSRA GVLsgvGKLv
    101 RQgAKFgTRA VPyVGTALLA HDVYETFKED IQAGCRyDP ETDKfVKGyE
    151 YANCLWYEDe RRINRTYGCY GVDSSIMRLM PDRSRfPEVK QLMESQMyRL
    201 ARPFWNWRKE ELNKLSSLDW NNFVLNRCtF DWNGGgCAVN KGDDfFRAGAS
45 251 FSLGRNPkYK EEMDAKKPEE ILSLKVDADP DKYIEATGYp GYSEKVEVAP
    301 GTKVNMGPVT DRNGNPVQVA ATfGRDAQGN TTADVQVIpR PDLTPASAEa
    351 PHAQPLPEVS PAENPANNPd PDENPGTRPN PEPDPDLNPd ANPDtDGQPG
    401 TSPDSPAVPD RPNGRHRKER KEGEDGGLSC DYfPEILACQ EMGKPSDRMF
    451 HDISIPQVTD DKTWSSHNF LPSNGVCPQPK TFHVfGRQYR ASYEPLCVfA
50 501 EKIRfAVILLa FIIMSAfVfE GS LGGE*

```

After further analysis, the following gonococcal DNA sequence <SEQ ID 275> was identified:

```

55 1  ATGGTCACAA AACATACAA TTTGAATTTT GCGAAATTGT CGATAATTGC
    51  AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAATGCA GTAAAAATAT
    101 CTGAAACTCT TTCGGTTGAT ACCGGACAAG GCGCGAAAGT TCATAAGTTC
    151 GTTCCTAAAT CAAGTAATAT TTATTCATCT GATTTAACAA AAGCGGTAGA
    201 TTTAACGCAT ATCCCCACGG GCGCAAAAGC CCGAATCAAC GCCAAATAA
    251 CCGCCAGCGT ATCCCGCGCC GGCGTATTGT CGGGGgTCGG CAAACTTGTC
    301 CGCCAAGGCG CGAAATTCGG CACAAGGGCG GTTCCCTATG TCGGAACAGC
    351 CCTTTTAGCC CACGACGTAT ACGAAACTTT CAAAGAAGAC ATACAGGCAC
60 401 GAGGCTGCCG ATACGATCCC GAAACCGACA AATTT

```

This corresponds to the amino acid sequence <SEQ ID 276; ORF72ng-1>:

```

1  MVTKHTNLNF AKLSIIAILM MYSFEANANA VKISETLSVD TGQGAHVHKF
51  VPKSSNIYSS DLTKAVDLTH IPTGAKARIN AKITASVSRA GVLGSGVGLV
101 RQGAKEGTRA VPYVGTALLA HDVYETFKED IQARGCRYDP ETDKF

```

5 ORF72ng-1 and ORF721-1 show 89.7% identity in 145 aa overlap:

```

10  orf72ng-1.pe 10 20 30 40 50 60
    orf72-1      10 20 30 40 50 60
    orf72ng-1.pe 70 80 90 100 110 120
    orf72-1      70 80 90 100 110 120
15  orf72ng-1.pe 130 140
    orf72-1      130 140 150
20  orf72ng-1.pe HDVYETFKEDIQARGCRYDPETDKF
    orf72-1      HDVYETFKEDIQARGQYDPETDKFAKVSXG

```

Based on this analysis, including the presence of a putative leader sequence and transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 33

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 277>:

```

30  1  ATGAGATTTT TCGGTATCGG TTTTGTGGTG CTGCTGTTTT TGGAGATTAT
    51  GTCGATTGTG TGGGTTGCCG ATTGGCTGGG CGGCGGCTGG ACGTTGTTTT
    101  TGATGGCGGC AGGTTTGGCC GCCGGCGTGC TGATGCTCAG GCAAACCGGG
    151  GCTGACCGGT CTTTATTGGG CGGGCGCGGC AATGAGAAGC GCGGGGAAGG
    201  TATCCGTTTA TCAGATGTTG TGGCCTATC..

```

35 This corresponds to the amino acid sequence <SEQ ID 278; ORF73>:

```

1  MRFFGIGFLV LLFLEIMSIV WVADWLGGGW TLFLMAAGFA AGVLMRLQTG
51  LTGLLLAGAA MRSGGKVSIV QMLWPI..

```

Further work revealed the complete nucleotide sequence <SEQ ID 279>:

```

40  1  ATGAGATTTT TCGGTATCGG TTTTGTGGTG CTGCTGTTTT TGGAGATTAT
    51  GTCGATTGTG TGGGTTGCCG ATTGGCTGGG CGGCGGCTGG ACGTTGTTTT
    101  TGATGGCGGC AGGTTTGGCC GCCGGCGTGC TGATGCTCAG GCATACGGGG
    151  CTGTCCGGTC TTTTATTGGC GGGCGCGGCA ATGAGAAGCG GCGGGAGGGT
    201  ATCCGTTTAT CAGATGTTGT GGCCTATCCG TTATACGGTG GCGGCTGTGT
    251  GTCTGATGAG TCCGGGATTC GTATCCTCGG TGTGGCGGT ATTGCTGCTG
45  301  CTGCCGTTTA AGGGAGGGGC AGTGTGTCAG GCAGGAGGTG CGGAAAATTT
    351  TTTCAACATG AACCAATCGG GCAGAAAAGA GGGCTTTTCC CGCGATGACG
    401  ATATTATCGA GGGAGAATAT ACGGTTGAAG AGCCTTACGG CGGCAATCGT
    451  TCCCGAAACG CCATCGAACA CAAAAAAGAC GAATAA

```

This corresponds to the amino acid sequence <SEQ ID 280; ORF73-1>:

```

50  1  MRFFGIGFLV LLFLEIMSIV WVADWLGGGW TLFLMAAGFA AGVLMRLRHTG
    51  LSGLLLAGAA MRSGGRVSIV QMLWPIRYTV AAVCLMSPGF VSSVLAVLLL
    101  LPFKGGAVLQ AGGAENFFNM NQSGRKEGFS RDDDIIEGEY TVEEYGGNR

```

151 SRNAIEHKKD E\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF73 shows 90.8% identity over a 76aa overlap with an ORF (ORF73a) from strain A of *N.*

5 *meningitidis*:

```

      10      20      30      40      50      60
orf73.pep  MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLFLMAAGFAAGVLMRLRQTGLTGLLLAGAA
          |||||:|||||:|||||:|||||:|||||:|||||
orf73a     MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLFLMAATFAAGVVMRLRHTGLSGLLLAGAA
      10      20      30      40      50      60

      70
orf73.pep  MRSGGKVSQMLWPI
          |||||:|||||
orf73a     MRSGGRVSVYXMLWIRYTVAAVCXMSPGFVSSVXAVLLXLPFKGGAVLQAGGAENFFNM
15

```

The complete length ORF73a nucleotide sequence <SEQ ID 281> is:

```

1  ATGAGATTTT TCGGTATCGG TTTTGTGGTG CTGCTGTTTT TGGAGATTAT
51  GTCGATTGTG TGGGTGCGG ATTGGTTGGG CGGCGGTTGG ACGCTGTTTC
101 TAATGGCGGC AACCTTTGCC GCCGGCGTGG TGATGCTCAG GCATACGGGG
20  CTGTCCGGTC TTTTATTGGC GGGCGCGGCA ATGAGAAGCG GCGGGAGGGT
201 ATCCGTTTAT CANATGTTGT GGCNTATCCG TTATACGGTG GCGGCGGTGT
251 GTCNGATGAG TCCGGGATTC GTATCCTCGG TGTNGGCGGT ATTGCTGNTG
301 CTNCCGTTTA AGGGAGGTGC AGTGTTCAG GCAGGAGGTG CGGAAAATT
25  TTTCAACATG AACCANTCGG GCAGAAAAGA NGGCNTTCC CGCGATGACG
401 ATATTATCGA GGGGGAATAT ACGGTTGAAG ANCCTACGG CGGCANTCGT
451 TTCCGAAACG CCNTNGAACA CAAAAAGAC GAATAA

```

This encodes a protein having amino acid sequence <SEQ ID 282>:

```

1  MRFFGIGFLV LLFLEIMSIW VADWLGGGW TLFLMAATFA AGVVMRLRHTG
51  LSGLLLAGAA MRSGGRVSVY XMLWIRYTV AAVCXMSPGF VSSVXAVLLX
30  101 LPFKGGAVLQ AGGAENFFNM NXSGRKXGXS RDDDIIEGEY TVEXPYGGXR
151 FRNAXEHKKD E*

```

ORF73a and ORF73-1 show 91.3% identity in 161 aa overlap

```

      10      20      30      40      50      60
orf73a.pep MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLFLMAATFAAGVVMRLRHTGLSGLLLAGAA
35  |||||:|||||:|||||:|||||:|||||:|||||
orf73-1    MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLFLMAAGFAAGVLMRLRHTGLSGLLLAGAA
      10      20      30      40      50      60

      70      80      90      100     110     120
orf73a.pep MRSGGRVSVYXMLWIRYTVAAVCXMSPGFVSSVXAVLLXLPFKGGAVLQAGGAENFFNM
40  |||||:|||||:|||||:|||||:|||||:|||||
orf73-1    MRSGGRVSVYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLPFKGGAVLQAGGAENFFNM
      70      80      90      100     110     120

      130     140     150     160
orf73a.pep NXSGRKXGXS RDDDIIEGEYTVEXPYGGXRFRNAXEHKKDEX
45  |||||:|||||:|||||:|||||:|||||:|||||
orf73-1    NQSGRKEGFSRDDDIIEGEYTVVEPYGGNRSRNAIEHKKDEX
      130     140     150     160
50

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF73 shows 92.1% identity over a 76aa overlap with a predicted ORF (ORF73.ng) from *N.*

*gonorrhoeae*:

```

55  orf73.pep  MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLFLMAAGFAAGVLMRLRQTGLTGLLLAGAA  60
          |||||:|||||:|||||:|||||:|||||:|||||

```

```

      orf73ng      MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLMAATFAAGVLMRLRHTGLSGLLLAGAA 60
      orf73.pep    MRSGBKSVYQMLWPI 76
      orf73ng      VKSSGKSVYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFGKGAVLQAGGAENFFNM 120

```

The complete length ORF73ng nucleotide sequence <SEQ ID 283> is:

```

      1  ATGAGATTTT TCGGTATCGG TTTTGTGGT CTGCTGTTTT TGGAAATTAT
      51  GTCGATTGTG TGGGTTGCCG ATTGGCTGGG CGGCGGTTGG AcgctGTTC
     101  TAATGGCGGC AACCTTTGCC GCCGTGTGTC TGATGCTCAG GCATAcggGG
     151  CTGTCCGGTC TTTATTGGC TGGCGCGCGC GTAAAAagta gtgGGAAGGT
     201  ATCTGTTTAT CagatgtTGT GGCCTATCCG TTATAcggtg gcggcggtgT
     251  GTCTGatgag tCcgGATTC GTATCCTccg tgttgGCGGT ATTGCTGCTG
     301  CTGCcgttta aggGaggGgc agtgttgGag gcaggaggtg cggaaaATT
     351  TTTCAACATg aaCcaatcg gcagaaAaga gggatttttc cacgatgacg
     401  atattatcga gggagaatat acggttgaaa aacctgacgg cggcaatcgt
     451  tcccgAAcgc ccatcgaaac cgaaaAagac gaataA

```

This encodes a protein having amino acid sequence <SEQ ID 284>:

```

      1  MRFFGIGFLV LLFLEIMSIW VVADWLGGGW TFLMAATFA AGVLMRLRHTG
     51  LSGLLLAGAA VKSSGKSVY QMLWPIRYTV AAVCLMSPGF VSSVLAVLLL
    101  LPFKGAVALQ AGGAENFFNM NQSGRKEGFF HDDDIIEGEY TVEKPDGGNR
    151  SRNAIEHEKD E*

```

ORF73ng and ORG73-1 show 93.8% identity in 161 aa overlap

```

      10      20      30      40      50      60
    orf73-1.pep MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLMAAGFAAGVLMRLRHTGLSGLLLAGAA
    orf73ng      MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLMAATFAAGVLMRLRHTGLSGLLLAGAA
      10      20      30      40      50      60
      70      80      90     100     110     120
    orf73-1.pep MRSGRVSVYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFGKGAVALQAGGAENFFNM
    orf73ng      VKSSGKSVYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFGKGAVALQAGGAENFFNM
      70      80      90     100     110     120
      130     140     150     160
    orf73-1.pep NQSGRKEGFSRDDDIIEGEYTVVEEPYGGNRSRNAIEHKKDEX
    orf73ng      NQSGRKEGFFHDDDIIEGEYTVVEEPYGGNRSRNAIEHEKDEX
      130     140     150     160

```

Based on this analysis, including the presence of a putative leader sequence and putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 34

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 285>:

```

      1  ATGTTTGTTC TTCAGACGGC ATTCTT.ATG TTTCAGAAAC ATTTGCAGAA
     51  AGCCTCCGAC AGCGTCGTCG GAGGGACATT ATACGTGGTT GCCACGCCCCA
    101  TCGGCAATTT GCGGACATT ACCCTGCGCG CTTTGGCGGT ATTGCAAAAG
     151  GCG..... .GCGA AGACACGCGC GTTACCGCAC AGCTTTTGAG
     201  CGCGTACGGC ATTCAGGGCA AACTCGTCAG TGTGCGCGAA CACAACGAAC
     251  GGCAGATGGC GGACAAGATT GTCGGCTATC TTTCAGACGG CATGGTTGTG
     301  GCACAGGTTT CCGATGCGGG TACGCCGGCC GTGTGCGACC CGGCGCGGAA
     351  ACTCGCCCGC CGCGTGCCTG AGGCCGGGTT TAAAGTCGTT CCCGTCGTGG
     401  GCGCAAC.GC GGTGATGGCG GCTTTGAGCG TGGCCGGTGT GGAAGGATCC
     451  GATTTTATT TCAACGGTTT TGTACCGCCG AAATCGGGAG AACGCAGGAA

```

501 ACTGTTTGCC AAATGGGTGC GGGCGGCGTT TCCTATCGTC ATGTTTGAAA  
 551 CGCCGCACCG CATCGGTGCA GCGCTTGCCG ATATGGCGGA ACTGTTCCCC  
 601 GAACGCCGAT TAATGCTGCG GCGCGAAATT ACGAAAACGT TTGAAACGTT  
 651 CTTAAGCGGC ACGGTTGGGG AAATTCAGAC GGCATTGTCT GCCGACGGCG  
 701 ACCAATCGCG CGGCGAGATG GTGTGGGTGC TTTATCCGCG GCAGGATGAA  
 751 AAACACGAAG GCTTGTCCGA GTCCGCGCAA AACATCATGA AAATCCTCAC  
 801 AGCCGAGCTG CCGACCAAAC AGGCGGCGGA GCTTGCTGCC AAAATCACGG  
 851 GCGAGGGAAG GAAAGCTTTG TACGAT..

This corresponds to the amino acid sequence <SEQ ID 286; ORF75>:

10 1 MFVFQTAFXM FQKHLQKASD SVVGGTLYV ATPIGNLADI TLRALAVLQK  
 51 A...AEDTR VTAQLLSAYG IQGKLVSVRE HNERQMADKI VGYLSDGMVV  
 101 AQVSDAGTPA VCDPGAKLAR RVREAGFKVV PVVGAXAVMA ALSVAGVEGS  
 151 DFYFNGFVPP KSGERRKLFA KWRRAAFPIV MFETPHRIGA ALADMAELFP  
 201 ERRMLLAREI TKTFETFLSG TVGEIQTALS ADGDQSRGEM VLVLYPAQDE  
 15 251 KHEGLSESAQ NIMKILTAEI PTKQAAELAA KITGEGKKAL YD..

Further work revealed the complete nucleotide sequence <SEQ ID 287>:

1 ATGTTTCAGA AACATTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC  
 51 ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCGGAC ATTACCTGCG  
 101 GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG  
 151 CGCGTTACCG CACAGCTTTT GAGCGCGTAC GGCATTACAG GCAAACCTCGT  
 201 CAGTGTGCGC GAACACAACG AACGGCAGAT GCGGACAAAG ATTGTGCGCT  
 251 ATCTTTTACA CGGCATGGTT GTGGCACAGG TTTCCGATGC GGGTACGCCG  
 301 GCCGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GTGAGGCCCG  
 351 GTTTAAAGTC GTTCCCGTCG TGGGCGCAAG CGCGGTGATG CGGGCTTTGA  
 25 401 GCGTGGCCGG TGTGGAAGGA TCCGATTTTT ATTTCAACGG TTTTGTACCG  
 451 CCGAAATCGG GAGAACGCAG GAAACTGTTT GCCAAATGGG TCGGGGCGGC  
 501 GTTTCCTATC GTCATGTTTG AAACGCCGCA CCGCATCGGT GCGACGCTTG  
 551 CCGATATGGC GGAAGTGTTC CCCGAACGCC GATTAATGCT GCGCGCGGAA  
 601 ATTACGAAAA CGTTTGAAGC GTTCTTAAGC GGCACGGTGT GGGAAATTCA  
 30 651 GACGGCATTG TCTGCCGACG GCAACCAATC GCGCGGCGAG ATGGTGTGTTG  
 701 TGCTTTATCC GCGCGAGGAT GAAAAACACG AAGGCTTGTC CGAGTCCGCG  
 751 CAAAACATCA TGAAATCCT CACAGCCGAG CTGCCGACCA AACAGGCGGC  
 801 GGAGCTTGCT GCCAAATCA CGGGCGAGGG AAAGAAAGCT TTGTACGATC  
 851 TGGCTCTGTC TTGAAAAAAC AAATAG

35 This corresponds to the amino acid sequence <SEQ ID 288; ORF75-1>:

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT  
 51 RVTAQLLSAY GIQKLVSVR EHNERQMA DKIVGYLSDGMV VAQVSDAGTP  
 101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVEG SDFYFNGFVP  
 40 151 PKSGERRKLF AKWVRAAFPI VMFETPHRIG ATLADMAELF PERRMLLARE  
 201 ITKTETFLS GTVGEIQTAL SADGNQSRGE MVLVLYPAQD EKHEGLSESA  
 251 QNIMKILTAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF75 shows 95.8% identity over a 283aa overlap with an ORF (ORF75a) from strain A of *N. meningitidis*:

45

orf75.pep	10	20	30	40	50	60
	MFVFQTAFXM	FQKHLQKAS	DSVVGGTLYV	VATPIGNLAD	ITLRALAVLQ	KAXXXXAEDTR
orf75a						
	10	20	30	40	50	
	MFQKHLQKAS	DSVVGGTLYV	VATPIGNLAD	ITLRALAVLQ	KADIICAEDTR	

50

orf75.pep	70	80	90	100	110	120
	VTAQLLSAYG	IQGKLVSVRE	HNERQMA DK	IVGYLSDGMV	VAQVSDAGT	PAVCDPGAKLAR
orf75a						
	70	80	90	100	110	
	VTAQLLSAYG	IQGKLVSVRE	HNERQMA DK	IVGYLSDGMV	VAQVSDAGT	PAVCDPGAKLAR

55

orf75.pep	130	140	150	160	170	180
	RVREAGFKVV	PVVGAXAVMA	AALSVAGVEG	SDFYFNGFVP	PKSGERRKLF	AKWVRAAFPIV

	1	ATGTTTCAGA	AACATTTCGA	GAAAGCCTCC	GACAGCGTCG	TCGGAGGGAC
20	51	ATTATACGTG	GTTGCCACGC	CCATCGGCAA	TTTGGCGGAC	ATTACCCTGC
	101	GCGCTTTGGC	GGTATTGCAA	AAGGCGGACA	TCATCTGTGC	CGAAGACACG
	151	CGCGTTACCG	CGCAGCTTTT	GAGCGCGTAC	GGCATTTCAGG	CGAAACTCGT
	201	CAGCGTGCGC	GAACACAACG	AACGGCAGAT	GGCGGACAAG	ATTGTCTGGCT
	251	ATCTTTCAGA	CGGCATGGTT	GTGGCACAGG	TTTCCGATGC	GGGTACGCCG
25	301	GCCGTGTGCG	ACCCGGGCGC	GAAACTCGCC	CGCCGCGTGC	GTGAGGTCGG
	351	GTTTAAAGTT	GTCCCTGTTG	TCGGCGCAAG	CGCGGTGATG	CGCGCTTTGA
	401	GTGTGGCTGG	TGTGGCGGGA	TCCGATTTTT	ATTTCAACGG	TTTTGTACCG
	451	CGGAAATCGG	GCGAACGTAG	GAAATTGTTT	GCCAAATGGG	TGCGGGTGGC
	501	GTTTCCCGTC	GTGATGTTTG	AAACGCCGCA	CCGATCTCGG	GCGACGCTTG
30	551	CCGATATGGC	GGAACTGTTT	CCCGAACGCC	GATTAATGCT	GGCGCGCGAA
	601	ATCACGAAAA	CGTTTGAAC	GTTCTTAAAG	GGACCGGTTG	GGGAAATTCA
	651	GACGGCATTG	GCGGCGGACG	GCAACCAATC	GCGCGGCGAG	ATGGTGTGTTG
	701	TGCTTTATCC	GGCGCAGGAT	GAAAAACAG	AAGGCTTGTC	CGAGTCCCGC
	751	CAAAACATCA	TGAAAACTCT	CACAGCCGAG	CTGCCGACCA	AACAGCGCGC
35	801	GGAGCTTGCC	GCCAAAATCA	CGGGCGGAGG	AAAAAAAGCT	TTGTACGATC
	851	TGGCACTGTC	TTGGAAAAAC	AAATGA		

	1	MFQKHLQKAS	DSVVGTLVY	VATPIGNLAD	ITLRALAVLQ	KADIICAEDT
	51	RVTAQQLSAY	GIQGLVSVR	EHNERQMA DK	IVGYLSDGMV	VAQVSDAGTP
40	101	AVCDPGAKLA	RRVREVGFKV	VPVVGASAVM	AALSVAGVAF	SDFYFNGFVP
	151	PKSGERRKLF	AKWVRVAFPV	VMFETPHRIG	ATLADMAELF	PERRRLMLARE
	201	ITKTFTETFS	GTVGEIQ TAL	AADGNQSRGE	MVLVLYPAQD	EKHEGLSESA
	251	QNIMKILTAE	LPTKQAAELA	AKITGEGKKA	LYDLALSWK N	K*

45			10	20	30	40	50	60
	orf75a.pep	MFQKHLQKASDSVVG	TLYV	VATPI	GNLADIT	LRALAV	LQKADI	ICAEDTRVTAQLLSAY
	orf75-1	MFQKHLQKASDSVVG	TLYV	VATPI	GNLADIT	LRALAV	LQKADI	ICAEDTRVTAQLLSAY
		10	20	30	40	50	60	
50			70	80	90	100	110	120
	orf75a.pep	GIQGKLVSVREHNERQ	MADKIV	GYLSDG	MVVAQV	SDAGT	PAVCDP	GAKLARRVREVGFKV
	orf75-1	GIQGKLVSVREHNERQ	MADKIV	GYLSDG	MVVAQV	SDAGT	PAVCDP	GAKLARRVREAGFKV
55			70	80	90	100	110	120
	orf75a.pep	VPVVGASAVMAALSV	AGVAGS	DFYFNG	FVPPKSG	ERRKLF	AKWVRV	AFPPVMFETPHRIG
	orf75-1	VPVVGASAVMAALSV	AGVAGS	DFYFNG	FVPPKSG	ERRKLF	AKWVRV	AFPIVMFETPHRIG
60			130	140	150	160	170	180
	orf75a.pep	ATLADMAELFPERR	LMLARE	ITKT	FETFLS	GTVGEI	QTALA	ADGNQSRGEMVLVLYPAQD
	orf75-1	ATLADMAELFPERR	LMLARE	ITKT	FETFLS	GTVGEI	QTALA	ADGNQSRGEMVLVLYPAQD
65			190	200	210	220	230	240
	orf75a.pep m	ATLADMAELFPERR	LMLARE	ITKT	FETFLS	GTVGEI	QTALA	ADGNQSRGEMVLVLYPAQD

-298-

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

```

601 AACAAAcgca cgctcgaATT GCGCCACGGC TACCGTTACA GCGGcacgcC
651 CGGacGCGCc gactaCAATC AGGTTtcctt cCAAAAcTc aacctgATta
701 TCAGCACCCAC GCCCAAcTT ATCGaccCGG TTCCcACCG CCGCACcATT
751 tcgacCGCCC AAcTGATTGG CAGCAGCAAT CCGCAACATC AGGCAGAAATT
801 GATGTGGCGC ATCTCGCTGA CCGTCAGCGT CCTCCTGCTC TGCCTACTCG
851 CCGTGCCGCT TTCCTATTTT AACCCGCGCA GCGGACATAC CTACAATATC
901 TTGATTGCCA TCGGTTTGTT TTTAATTAC CAAAACGGGC TGACCCTGCT
951 TTTTGAAGCC GTGGAAGACG GCAAATCCA TTTTGGCTC GGACTGCTGC
1001 CTATGCACAT CATCATGTTT GTCATCGCAA TCGTACTTCT GCGCGTCCGC
1051 AGTATGCCCA GCCAGCCCTT CTGGCAGGCG GTTGGCAAAA GTCTGACATT
1101 GAAAGgcgGA AAATGA

```

This corresponds to the amino acid sequence <SEQ ID 506; ORF101ng-1>:

15  
20  
25  
30  
35  
40  
45  
50  
55  
60

```

1 MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GRVAIDAVLA
51 LVGFVWIGMT PLLLVLTAFI STLTVLTRYW RDSEMSVWLS CGLALKQWIR
101 PVMQFAVPEFA ILIAVMQLWV IPWAE LRSRE YAEILKQKQE LSLVEAGEFN
151 NLGKRNGRVY FVETFDTESG IMKNLFLREQ DKNGGDNIIF AKEGNFSLKD
201 NKRTLELRHG YRYSCTPGRA DYNQVSFOKL NLIISTTPKL IDPVSHRRTI
251 STAQLIGSSN PQHQAE LMRW ISLTVSVLLL CLLAVPLSYF NPRSGHTYNI
301 LIAIGLFLIY QNGLTLLFEA VEDGKIHFWL GLLPMHIIIMF VIAIVLRLVR
351 SMPSQPFWQA VGKSLTLKGG K*

```

ORF101ng-1 and ORF101-1 show 97.6% identity in 371 aa overlap:

25  
30  
35  
40  
45  
50  
55  
60

```

      10      20      30      40      50      60
orf101-1.pep MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGRVAIDAVLALVGFWIGMT
orf101ng-1    MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGRVAIDAVLALVGFWIGMT
      10      20      30      40      50      60

      70      80      90     100     110     120
orf101-1.pep PLLLVLTAFISTLTVLTRYWRDSEMSVWLS CGLALKQWIRPVMQFAVPEFAVLVAVMQLWV
orf101ng-1    PLLLVLTAFISTLTVLTRYWRDSEMSVWLS CGLALKQWIRPVMQFAVPEFAILIAVMQLWV
      70      80      90     100     110     120

      130     140     150     160     170     180
orf101-1.pep IPWAE LRSREYAEILKQKQELSLVEAGEFN LGKRNGRVYFVETFDTESGIMKNLFLREQ
orf101ng-1    IPWAE LRSREYAEILKQKQELSLVEAGEFN LGKRNGRVYFVETFDTESGIMKNLFLREQ
      130     140     150     160     170     180

      190     200     210     220     230     240
orf101-1.pep DKNGGDNIIFAKEGNFSLNDNKRTLELRHGYRYSCTPGRADYNQVSFOKL NLIISTTPKL
orf101ng-1    DKNGGDNIIFAKEGNFSLNDNKRTLELRHGYRYSCTPGRADYNQVSFOKL NLIISTTPKL
      190     200     210     220     230     240

      250     260     270     280     290     300
orf101-1.pep IDPVSHRRTIPTAQLIGSSNPQHQAELMRWISLTVSVLLLCLLAVPLSYFNPRSGHTYNI
orf101ng-1    IDPVSHRRTISTAQLIGSSNPQHQAELMRWISLTVSVLLLCLLAVPLSYFNPRSGHTYNI
      250     260     270     280     290     300

      310     320     330     340     350     360
orf101-1.pep LIAIGLFLIYQNGLTLLFEAVEDGKIHFWLGLLPMHIIIMFAVALILLVRSMPSQPFWQA
orf101ng-1    LIAIGLFLIYQNGLTLLFEAVEDGKIHFWLGLLPMHIIIMFVIAIVLRLVRSMPSQPFWQA
      310     320     330     340     350     360

      370
orf101-1.pep VGKSLTLKGGKX
orf101ng-1    VGKSLTLKGGKX
      370

```

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is

predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 60

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 507>:

```

5      1  ..GGTGGTGGTT TTATCAATGC TTCCTGTGCC ACTTTGACGA CAGCCAAACC
      51  GCAATATCAA GCAGGAGACC TTAGCGCTTT TAAGATAAGG CAAGGCAATG
     101  TTGTAATCGC CGGACACGGT TTGGATGCAC GTGATACCGA TTACACACGT
     151  ATTCTCAGTT ATCATTCCAA AATCGATGCA CCCGTATGGG GACAAGATGT
     201  TCGTGTCTGC GCGGGACAAA ACGATGTGGC CGCAACAGGT GATGCACATT
    10  251  CGCCTATTCT CAATAATGCT GCTGCCAATA CGTCAAACAA TACAGCCAAC
     301  AACGGCACAC ATATCCCTTT ATTTGCGATT GATACAGGCA AATTAGGAGG
     351  TAT.GTATGC CAACAAAATC ACCTTGATCA GTACGGTCGA GCAAGCAGGC
     401  ATTCGTAA
  
```

This corresponds to the amino acid sequence <SEQ ID 508; ORF113>:

```

15      1  ..GGGFINASCA TLTTAKPQYQ AGDLSAFKIR QGNVVIAGHG LDARDTDYTR
      51  ILSYHSKIDA PVWQDVRVV AGQNDVAATG DAHSPIILNNA AANTSNNNTAN
     101  NGTHIPLFAI DTGKLGGXVC QQNHLDDQYGR ASRHS*
  
```

Computer analysis of this amino acid sequence gave the following results:

Homology with with pspA putative secreted protein of *N.meningitidis* (accession AF030941)

20 ORF and pspA show 44% aa identity in 179aa overlap:

```

      orf113  GGGFINASCATLTTAKPQYQAGDLSAFKIRQGNVVIAGHGLDARDTDYTRILSYHSKIDA 60
      pspA    GGG INA+ TLT+ P G+L+ F + G VVI G GLD D DYTRILS ++I+A
      pspA    GGGLINAASVTLTSGVPVLNNGNLTFGFDVSSGKVIVIGKGLDTSADYTRILSRAEINA 256
    25      orf113  PVWQDVRVVAGQNDVAATGDAHSPIILXXXXXXXXXXXXXXXXGTHIPLFAIDTGKLGGMYA 120
      pspA    VWG+DV+VV+G+N + G + P AIDT LGGMYA
      pspA    GVWGDVKVVS GKNKLD F D G-----SLAKTASAPSSSDSVTPTVAIDTATLGGMYA 307
    30      orf113  NKITLISTVEQAGIRNQGWFAAGNVAEGLVNTGMIAATGENHAVSLHARNVHN 179
      pspA    +KITLIST A IRN+G+ FA+ G V ++A+GKL N+G I A +++ A+ V N
      pspA    DKITLISTDNGAVIRNKGRIFAATGGVTLSDGKLSNNGSIDAA----EITISAQTVDN 362
  
```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF113 shows 86.5% identity in 52aa overlap at the N- terminal part and 94.1% identity in 17aa

35 overlap at the C-terminal part with a predicted ORF (ORF113ng) from *N. gonorrhoeae*:

```

      orf113  GGGFINASCATLTTAKPQYQAGDLSAFKIR 30
      orf113ng SHPSQLNGYIEVGGRRAEVVIANPAGIAVNGGGFINASRATLTGQPQYQAGDFSGFKIR 224
    40      orf113  QGNVVIAGHGLDARDTDYTRILSYHSKIDAPVWQDVRVVAGQNDVAATGDAHSPIILNNA 90
      orf113ng QGNAVIAGHGLDARDTDFTRILVCQQNHLDDQYGRSTRHS 263
    45      orf113  IDTGKLGGXVCQQNHLDDQYGRASRHS 135
      orf113ng DFGFKIRQGNVVIAGHGLDARDTDFTRILVCQQNHLDDQYGRSTRHS 263
  
```

The complete length ORF113ng nucleotide sequence <SEQ ID 509> is predicted to encode a protein having amino acid sequence <SEQ ID 510>:

```

50      1  MNKTLRYVIF NRKRGAVVAV AETTKREGKS CADSGSGSVY VKSVSFIFTH
      51  SKAFCFSA LG FSLCLALGTV NIAFADGIIT DKAAPKTQQA TILQTGNIGP
  
```



101 QVNIQTPTSA GVSVNQYAQF DVGNRGAILN NSRSNTQTQL GGWIQGNPWL  
 151 TRGEARVVVN QINSSHPSQL NGYIEVGGRR AEVVIANPAG IAVNGGGFIN  
 201 ASRATLTTGQ PQYQAGDFSG FKIRQGNAPI AGHGLDARDT DFTRILVCQQ  
 251 NHLDQYGRTS RHS\*

- 5 Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 61

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 511>:

```

10      1  ..TCAACGGGAC ATAGCGAACA AAATTACACT TTGCCGCGAG AAATCACACG
      51  CAACATTTCAT CTGGGTTTCAT TTGCCTATGA ATCGCATCGC AAAGCATTAA
      101  GCCATCATGC GCCCAGCCAA GGCACAGTGTG TGCCGCAAAG CAACGGTATT
      151  TCGCTACCCT ATACGTCCAA TTCTTTTACC CCATTACCCA GCAGCAGCTT
      201  ATACATTATC AATCCTGTCA ATAAAGGCTA TCTTGTGAA ACCGATCCAC
      251  GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GCTGGACAGC
      15  301  CTCAAACACTAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGGTTATTA
      351  CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC
      401  GTTTAGACGG TTATCAAAAC GACGAAGAAC AATTAAAGC CTTAATGGAT
      451  AATGGCGCGA CTGCGGCACG TTcGATGAAT CTCAGCGTTG GCATTGCATT
      501  AAGTGCCGAG CAAGTAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC
      20  551  AAAAAGAAGT TAAGCTTCCT GATGGCGGCA CACAAACCGT ATTGGTGCCA
      601  CAGGTTTATG TACGCGTTAA AAATGGCGAC ATAGACGGTA AAGGTGCATT
      651  GTTGTACAGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT
      701  CAGGCACGAT TGCAGGgCGC AATGCGCTTA TTATCAATAC CGATACGCTA
      751  GACAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC
      25  801  ACAAGACATC AATAATATTG GCGGCATGCT TTCTGCCGAA CAGACATTAT
      851  GCTCAACGC AGGCAACAAC ATCAACAGCC AAAGCACCAC CGCCAGCAGT
      901  CAAAATACAC AAGGCAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA
      951  TATCACAGGC AAAGAAAAAG GTGTTT..
  
```

This corresponds to the amino acid sequence <SEQ ID 512; ORF115>:

```

30      1  ..STGHSEQNYT LPREITRNI LGSFAYESH KALSHHAPSQ GTELPQSNGI
      51  SLPYTSNSFT PLPSSSLYII NPVNKGYLE TDPRFANYRQ WLGS DYMLDS
      101  LKLDPNNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
      151  NGATAARSMN LSVGIALSAE QVAQLTSDIV WLVOKEVKLP DGGTQTVLVP
      201  QYVVRVKNKD IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
      35  251  DNIGGRIHAQ KSAVTATQDI NNIGGMLSAE QTLNAGNN INSQSTTASS
      301  QNTQGSSTYL DRMAGIYITG KEKGV..
  
```

Computer analysis of this amino acid sequence gave the following results:

Homology with the *pspA* putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF115 and *pspA* protein show 50% aa identity in 325aa overlap:

```

40  Orf115: 1  STGHSEQNYTLPREITRNI LGSFAYESH KALSHHAPSQGT ELPQSNGISLPYTSNSFT 60
      STG+S  Y  E++ +I +G AY+ +  +  P  +  NGI  +T
      pspA:  778 STGYRSRSPYEPAPEVS-SIRMGISAYKGYAPQQASDIPGTVVPVVAENG IHPFT----- 831

      Orf115: 61  PLPSSSLYIINPVNKGYLETDPRFANYRQWLGS DYMLDSLKLDPNNLHKRLGDGYEQR 120
      LP+SSL+ I P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYEQ+
      pspA:  832 -LPNSSLFAIAPNNKGYLIETDPAFTDYRKWLGS GYMLAALQQDPNHIHKRLGDGYEQK 890

      Orf115: 121 LINEQIAELTGHRRLDGYQNDDEEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIV 180
      L+NEQIA+LTG+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQVA+LTSDIV
      pspA:  891 LVNEQIAKLTGYRRLDGYTNDDEEQFKALMDNGITIAKELQLTPGIALSAEQVARLTSDIV 950

      Orf115: 181 WLVOKEVKLPDGGTQTVLVPQYVVRVKNKDIDGKGALLSGSNTQINVSGSLKN-SGTIAG 239
      WL  + V LPDG TQTVL P+VYVR + D++G+GALLSGS I SG+++N G IAG
      pspA:  951 WLENETVTLPDGTTQTVLKPVKYVRARPKDMNGQALLSGSVVDIG-SGAIENRGGLIAG 1009

55
  
```

10 ORF115 shows 91.9% identity over a 334aa overlap with a predicted ORF (ORF115ng) from *N.gonorrhoeae*:

	orf115.pep	STGHSEQNYTLPREITRNISLGSFAYESHK	31
		:    :	
15	orf115ng	NEQTFGEKKVFSENGKLHNYWRARRKGHDETGHRQNYTLPEEITRDISLGSFAYESHK	71
	orf115.pep	ALSHHAPSQGTLPQSN-----GISLPYTSNSFTPLPSSSLYIINPVNKGYLVT	81
		:                   :     :	
	orf115ng	ALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYIINPANKGYLVT	131
20	orf115.pep	DPRFANYRQWLGS DYMLDSLKLDPNNLHKRLGDGYEQRLINEQIAELTGHRRLDGYQND	141
	orf115ng	DPRFANYRQWLGS DYMLGSLKLDPNNLHKRLGDGYEQRLINEQIAELTGHRRLDGYQND	191
25	orf115.pep	EEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIVWL VQKEVKLPDGGTQTVLPQ	201
	orf115ng	EEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWL VQKEVKLPDGGTQTVLMPQ	251
	orf115.pep	VYVRVKNGDIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK	261
30	orf115ng	VYVRVKNGGIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK	311
	orf115.pep	SAVTATQDINNIGGMLSAEQTL L L L NAGNNINSQSTTASSQNTQGSSTYLDRMAGIYITGK	321
	orf115ng	SAVTATQDINNIGGILSAEQTL L L L NAGNNINNQSTAKSSQNAQGSSTYLDRMAGIYITGK	371
35	orf115.pep	EKGV	325
	orf115ng	EKGVLAQAQKDINIIAGQISNQSDOGOTRLOAGRDNLDTVOTGKYOEIHFADNHTIR	431

An ORF115ng nucleotide sequence <SEQ ID 513> was predicted to encode a protein having amino acid sequence <SEQ ID 514>:

	1	MLVQTEKDGL	HNEQTFGEKK	VFSENGKLHN	YWRARRKGHD	ETGHRQNYT
	51	LPEETTRDIS	LGSFAYESH	KALSRHAPSQ	GTELPQSNRD	NIRTAKSNGI
	101	SLPYTPNSFT	PLPGSSLYII	NPANKGYLVE	TDPFRANYRQ	WLGSIDYMLGS
45	151	LKLDPNNLHK	RLGDGYEOR	LINEQIAELT	GHRRLDGYQN	DEEQFKALMD
	201	NGATAARSMN	LSVGIALSAE	QAAQLTSDIV	WLWQKEVKLP	DGGTQTVLMP
	251	QVYVRVKNGG	IDKGALLSG	SNTQINVSGS	LKNSSTIAGR	NALINTDTL
	301	DNIGGRIHAQ	KSAVTATQDI	NNIGGILSAE	QTLLLNAGNN	INNQSSTAKS
	351	QNAQGSSTYL	DRMAGIYITG	KEKGVLAQA	KGDINIAGQ	ISNQSDDGGQ
50	401	RLQAGRDINL	DTVQTKGYQE	IHFADADNHTI	RGSTNEVGSS	IQTKGDVTL
	451	SGNNLNAKAA	EVGSAKGTLA	VYAKNDITIS	SGIHAGQVDD	ASKHTGRSGG
	501	GKLVITDKA	QSHHETAQSS	TFEKGQVVLQ	AGNADANILGS	NVISDNGTRI
	551	QAGNHVRIGT	TQTQSQSEY	HQTQKSGLMS	AGIGFTIGSK	TNTQENQSQS
	601	NEHTGSTVGS	LKGDTTIVAS	KHYEQTGSNV	SSPEGNLIS	TQSMIDGAAQ
	651	NQLNSKTTQT	YEQKGLTVAF	SSPVTDLAQ	AIATAHKA	QFDKAKTTAL
55	701	MPWRLPMOVG	RLFKQAKAPK	K*		

Further work revealed the following partial gonococcal DNA sequence <SEQ ID 515>:

60

1	TTGCTTGTGC	AAACAGAAAA	AGACGGTTTG	CATAACGAGC	AAACCTTTGG
51	CGAGAAAGAA	GTCTTCAGCG	AAATTGGTAA	GTTCGACAAC	TACTGGCGTG
101	CGCGTCGTAA	AGGCATGAT	GAAACAGGGC	ATCGTGAAAC	AAATTACT
151	TTGCCGGAGG	AAATCACACG	CGACATTTCA	CTGGGTTCAT	TTGCCTATGA
201	ATCGCATAGC	AAAGCATTAA	GCCGTCATGC	GCCGAGCCAA	GGCCTAGAT
251	TGCCACAAAG	TAACCCGGAT	AATATCCGTA	CTCCGAAAAG	CAACCGTATT

-302-

5 301 TCGCTACCCT ATACGCCCAA TTCTTTTACC CCATTACCCG GCAGCAGCTT  
 351 ATACATTATC AATCCTGCCA ATAAAGGCTA TCTTGTGAA ACCGATCCAC  
 401 GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GCTGGGCAGC  
 451 CTCAAACCTAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGGTTATTA  
 501 CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC  
 551 GTTTAGACGG TTATCAAAAC GACGAAGAAC AATTAAAGC CTTAATGGAT  
 601 AATGGCGCGA CTGCGGCACG TTCGATGAAT CTCAGCGTTG GCATTGCATT  
 651 AAGTGCCGAG CAAGCAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC  
 10 701 AAAAAGAAGT TAAACTTCCT GATGGCGGCA CACAAACCGT ATTGATGCCA  
 751 CAGGTTTATG TACGCGTTAA AAATGGCGGC ATAGACGGTA AAGGTGCATT  
 801 GTTGT CAGGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT  
 851 CAGGCACGAT TGCAGGGCGC AATGCGCTTA TTATCAATAC CGATACGCTA  
 901 GACAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC  
 951 ACAAGACATC AATAATATTG CGGCGATTCT TTCTGCCGAA CAGACATTAT  
 15 1001 TGCTCAATGC GGTAAACAAC ATCAACAACC AAAGCACGGC CAAGAGCAGT  
 1051 CAAATATGCAC AAGGTAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA  
 1101 TATCACAGGC AAAGAAAAAG GTGTTTTAGC AGCGCAGGCA GGCAAAGACA  
 1151 TCAACATCAT TGCCGTCGCA ATCAGCAATC AATCAGATCA AGGGCAAACC  
 1201 CGGCTGCAGG CAGGACGCGA CATTAACTTG GATACGGTAC AAACCGGCAA  
 20 1251 ATATCAAGAA ATCCATTTTG ATGCCGATAA CCATACCATC CGAGGTTCAA  
 1301 CGAACGAAGT CGGCAGCAGC ATTCAAACAA AAGGCGATGT TACCCTatTG  
 1351 TCAGGGAATA ATCTCAATGC CAAAGCTGCC GAAGTCGCA GCGCAAAGG  
 1401 CACACTTGCC GTGTATGCTA AAAATGACAT TACTATCAGC TCAGGCATCC  
 1451 ATGCCGCGCA AGTTGATGAT GCGTCCAAAC ATACAGGCAG AAGCGGCGGC  
 25 1501 GGTAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAACTGC  
 1551 TCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG  
 1601 ATGCCAACAT CCTTGGCAGT AATGTTATTT CCGATAATGG CACCCGGATT  
 1651 CAAGCAGGCA ATCATGTTCC CATTGGTACA ACCCAAATC AAAGCCAAAG  
 1701 CGAAACCTAT CATCAAACCC AAAAATCAGG ATTGATGAGT GCAGGTATCG  
 30 1751 GCTTCACTAT TGGCAGCAAG ACAAACACAC AAGAAAACCA ATCCCAAAGC  
 1801 AACGAACATA CAGGCAGTAC CGTAGGCAGC CTGAAAGGCG ATACCACCAT  
 1851 TGTGCAAGC AAACACTACG AACAAACCGG CAGCAACGTT TCCAGCCCTG  
 1901 AGGGCAACAA CCTTATCAGC ACGCAAAGTA TGGATATTGG CGCAGACAAA  
 1951 AACCAATTAA ACAGCAAAAC CACCCAAACC TACGAACAAA AAGGCTTAAC  
 35 2001 GGTGGCATT CAGTTCGCCCG TTACCGATTG GGCACAACAA GCGATTGCCG  
 2051 TAGCACACAA AGCAGCAAAAC AAGTCGGACA AAGCAAAAAC GACCGCGTTA  
 2101 ATGCCATGGC GGCTGCCAAT GCAGGTTGGC AGGCCTATCA AACAGGCAAA  
 2151 GGCGCACAAA ACTTAG

This corresponds to the amino acid sequence <SEQ ID 516; ORF115ng-1>:

40 1 LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT  
 51 LPEEITRDIS LGSFAYESHK KALSRHAPSQ GTELPQSNRD NIRTAKSNGI  
 101 SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPRFANYRQ WLGS DYMLGS  
 151 LKLDPNNLHK RLG DGYEYQR LINEQIAELT GHRRLDGYQN DEEQFKALMD  
 201 NGATAARSMN LSVGIALSAE QAAQLTSDIV WLQKEVKLP DGGTQTVLMP  
 45 251 QVYVRVKNNG IDGKALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL  
 301 DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLNLAGNN INNQSTAKSS  
 351 QNAQGSSTYL DRMAGIYITG KEKGVLAQA GKDINI IAGQ ISNQSDQOT  
 401 RLQAGRDLNL DTVQTKYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTL  
 451 SGNLNLAKAA EVGSAKGLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG  
 50 501 GNKLVIDDKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVISDNGTRI  
 551 QAGNHVRIGT TQTQSQSEY HQTKSGLMS AGIGFTIGSK TNTQENQSQS  
 601 NEHTGSTVGS LKGDITIVAS KHYEQTGSNV SSPEGNNLIS TQSMIDGAAQ  
 651 NQLNSKTTQT YEQKGLTVAF SSPVTDLAQQ AIAVAHKAAN KSDKAKTTAL  
 701 MPWRLEPMQVG RPIKQAKAHK T\*

55 This gonococcal protein (ORF115ng-1) shows 91.9% identity with ORF115 over 334aa:

20 30 40 50 60 70  
 orf115ng-1.p NEQTFGEKKVFSENGKLHNYWRARRKGHDETGHREQNYTLPEEITRDISLGSFAYESHK  
 orf115 ||| |||||:||||:|||||||  
 60 STGHSEQNYTLPREITRNISLGSFAYESHK  
 10 20 30  
 80 90 100 110 120 130  
 orf115ng-1.p ALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYIINPANKGYLVET  
 orf115 |||:||||||| ||||| |||||:|||||:|||||  
 65 ALSHHAPSQGTLPQSN-----GISLPYTPNSFTPLPSSLYIINPVNKGILVET  
 40 50 60 70 80

-303-

5 orf115ng-1.p 140 150 160 170 180 190  
 orf115 DPRFANYRQWLGS DYMLGSLKLPNNLHKRLGDGYEQRLINEQIAELTGHRRLDGYQND  
 90 100 110 120 130 140

10 orf115ng-1.p 200 210 220 230 240 250  
 orf115 EEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWLQKEVKLPDGGTQTVLMPQ  
 150 160 170 180 190 200

15 orf115ng-1.p 260 270 280 290 300 310  
 orf115 VYVRVKNGGIDGKGALLSGSNTQINVSGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQK  
 210 220 230 240 250 260

20 orf115ng-1.p 320 330 340 350 360 370  
 orf115 SAVTATQDINNIGGILSAEQTLNAGNNINNQSSTAKSSQNAQGSSTYLD RMAGIYITGK  
 270 280 290 300 310 320

25 orf115ng-1.p 380 390 400 410 420 430  
 orf115 EKGVLAAQAGKDINIIAGQISNQSDQGQTRLQAGRDINLDTVQTGKYQEIHFADNHTIR  
 EKG

In addition, it shows homology with a secreted *N.meningitidis* protein in the database:

30 gi|2623258 (AF030941) putative secreted protein [Neisseria meningitidis] Length  
 = 2273  
 Score = 604 bits (1541), Expect = e-172  
 Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)

35 Query: 1 LLVQTEKDGLHNEQTFGEKKVSENGKLHNYWRARRKGHDETGHREQNYTLPEEITRDIS 60  
 L+V T + L N++T G K + ++ G LH Y R +KG D TG+ Y E++ I  
 Sbjct: 739 LIVGTPESALDNDLGTGKTI-TDKGDLHRYHRHHKKGRDSTGYSRSPYEPAPSVS-SIR 796

40 Query: 61 LGSFAYESHKALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYII 120  
 +G AY+ + AP Q +++P + + NGI +T LP SSL+ I  
 Sbjct: 797 MGISAYKGY-----APQQASDIPGTV---VPVVAENGIHPTFT-----LPNSSLFAI 840

45 Query: 121 NPANKGYLVETDPRFANYRQWLGS DYMLGSLKLPNNLHKRLGDGYEQRLINEQIAELT 180  
 P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYEQ+L+NEQIA+LT  
 Sbjct: 841 APNNKGYLIETDPAFTDYRKWLGS GYMLAALQQDPNHIHKRLGDGYEQKLVNEQIAKLT 900

50 Query: 181 GHRRLDGYQNDDEEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWLQKEVKLP 240  
 G+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQ A+LTSDIVWL + V LP  
 Sbjct: 901 GYRRLDGYTNDEEQFKALMDNGITIAKELQTPGIALSAEQVARLTSDIVWLENETVTLP 960

55 Query: 241 DGGTQTVLMPQVYVRVKNGGIDGKGALLSGSNTQINVSGSLKN-SGTIAGRNALIINTDT 299  
 DG TQTVL P+VYVR + ++G+GALLSGS I SG+++N G IAGR ALI+N  
 Sbjct: 961 DGTQTVLKPQVYVRARPKDMNGQGALLSGSVVDIG-SGAIENRGGLIAGREALILNAQN 1019

60 Query: 300 LDNIGGRIHAQKSAVTATQDINNIGGILSAEQTLNAGNNINNQSSTAKSSQNAQGSSTY 359  
 + N+ G + + A DI N G I AE LLL A NNI ++S +S+QN QGS  
 Sbjct: 1020 IKNLQGD LQGNIFAAAGSDITNTGSI-GAENALLKASNNIESRSETRSNQNEQGSVRN 1078

65 Query: 360 LDRMAGIYITGKEKGVLAAGKDINIIAGQISNQSDQGQTRLQAGRDINLDTVQTGKYQ 419  
 + R+AGIY+TG++ G + AG +I + A +++NQS+ GQT L AG DI DT + Q  
 Sbjct: 1079 IGRVAGIYLTGRONGSVLLDAGNNIVLTASELTNQSEDGQTVLNAGGDIRSDTTGISRNQ 1138

70 Query: 420 EIHFDADNHTIRGSTNEVGSSIQTKGDVTLTSGNNLNAKAAEVGSAKGT LAVYAKNDITI 479  
 FD+DN+ IR NEVGSI+T+G+++L + ++ +AAEVGS +G L + A DI +  
 Sbjct: 1139 NTIFDSNYVIRKEQNEVGSTIRTRGNLSLNAKGDIRIRAAEVGSEQGRLKLAAGRDIKV 1198

Query: 480 SSGIHAGQVDDASKHTGRSGGKNLVTDKAQSHHETAQSSTFEGKQVVLQAGNDANILG 539  
 +G + +DA K+TGRSGGG K +T ++ + A S T +GK+++L +G D + G  
 Sbjct: 1199 EAGKAHTETEDALKYTGRSGGGIKQKMRHLKNQNGQAVSGTLDGKEIILVSGRDITVTG 1258

Query: 540 SNVISDNGTRIQAGNHVRIGTTQTQSQSEYHQTQKSGLM-SAGIGFTIGSKTNTQENQS 598  
 SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S  
 Sbjct: 1259 SNIIADNHTILSAKNNIVLKAAETRSRSAEMNKKEKSGLMGSGGIGFTAGSKKDTQTNRS 1318

5 Query: 599 QSNEHTGSTVGSGLKDDTTIVASKHYEQTGSNVSSPEGNNLISTQSMIDIGAAQNQLNSKTT 658  
 ++ HT S VGS L G+T I A KHY QTGS +SSP+G+ IS+ + I AAQN+ + ++  
 Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKHYTQTGSTISSPQGDVGISGKISIDAAQNRYSQESK 1378

10 Query: 659 QTYEQKGLTVAFFSPVTD 676  
 Q YEQKG+TVA S PV +  
 Sbjct: 1379 QVYEQKGVTVVAISVPVVN 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 62

15 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 517>:

```

1  ..TCAGGGAATA ACCTCAATGC CAAAGCTGCC GAAGTCAGCA GCGCAAACGG
51  TACACTCGCT GTGTCTGCCA ATAATGACAT CAACATCAGC GCAGGCATCA
101 ACACGACCCA TGTGATGAT GCGTCCAAAC ACACAGGCAG AAGCGGTGGT
151 GGCAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACCGC
20  CCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG
251 ATGCCAACAT CCTTGGCAGC AATGTTATTT CCGATAATGG CACCCAGATT
301 CAAGCAGGCA ATCATGTTTC CATTGGTACA ACCCAAACCTC AAAGCCAAAG
351 CGAAACCTAT CATCAAACCC AGAAATCAGG ATTGATGAGT GCAGGTATCG
401 GCTTCACTAT TGGCAGCAAG ACAAACACAC AAGAAAACCA ATCCCAAAGC
25  AACGAACATA CAGGCAGTAC CGTAGGCAGC TTGAAAGGCG ATACCACCAT
501 TGTTCAGGC AAACACTACG AACAAATCGG CAGTACCGTT TCCAGCCCGG
551 AAGGCAACAA TACCATCTAT GCCCAAAGCA TAGACATTCA AGCGGCACAC
601 AACAAATTAA ACAGTAATAC CACCCAAACC TATGAACAAA AAGG.CTAAC
651 GGTGGCATT CAGTTCGCCC TTACCGATT GGCACAACAA ...

```

30 This corresponds to the amino acid sequence <SEQ ID 518; ORF117>:

```

1  ..SGNNLNAAKAA EVSSANGTLA VSANNDINIS AGINTTHVDD ASKHTGRSGG
51  GNKLVIDDKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVISDNGTQI
101 QAGNHVRIGT TQTQSQSEY HQTQKSGLMS AGIGFTIGSK TNTQENQSQS
151 NEHTGSTVGS LKGDTTIVAG KHYEQIGSTV SSPEGNNTIY AQSIDIQAAH
35 201 NKLNSNTTQT YEQKXLTVAF SSPVTDLAQQ ...

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the pspA putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF117 and pspA protein show 45% aa identity in 224aa overlap:

40 Orf117: 4 NLNAKAAEVSSANGTLAVSANNDINISAGINTTHVDDASKHTGRSGGGNKLVIDDKAQSH 63  
 ++ +AAEV S G L ++A DI + AG T +DA K+TGRSGGG K +T ++  
 pspA: 1173 DIRIRAAEVGSEQRLKLAAGRIDKVEAGKAHTETEDALKYTGRSGGGIKQKMRHLKNQ 1232

45 Orf117: 64 HETAQSSTFEGKQVVLQAGNDANILGSNVISDNGTQIQAGNHVRIGTTQTQSQSEYHQT 123  
 + A S T +GK+++L +G D + GSN+I+DN T + A N++ + +T+S+S ++  
 pspA: 1233 NGQAVSGTLDGKEIILVSGRDITVTGSNIIADNHTILSAKNNIVLKAAETRSRSAEMNKK 1292

50 Orf117: 124 KSGLM-SAGIGFTIGSKTNTQENQSQSNEHTGSTVGSGLKDDTTIVAGKHYEQIGSTVSS 182  
 +KSGLM S GIGFT GSK +TQ N+S++ HT S VGS L G+T I AGKHY Q GST+SS  
 pspA: 1293 EKSGLMGSGGIGFTAGSKKDTQTNRSETVSHTESVVGSLNGNTLISAGKHYTQTGSTISS 1352

Orf117: 183 PEGNNTIYAQSIDIQAAHNKLNSNTTQTYEQKXLTVAFFSPVTD 226  
 P+G+ I + I I AA N+ + + Q YEQK +TVA S PV +  
 pspA: 1353 PQGDVGISGKISIDAAQNRYSQESKQVYEQKGVTVVAISVPVVN 1396

Homology with a predicted ORF from *N.gonorrhoeae*

ORF117 shows 90% identity over a 230aa overlap with a predicted ORF (ORF117ng) from *N.gonorrhoeae*:

5	orf117.pep	SGNNLNAAAEVSSANGTLAVSANNDINIS	30
	orf117ng	IHFDADNHTIRGSTNEVGSSIIQTKGDVTLTSGNNLNAAAEVGSAGKTLAVYAKNDITIS	480
	orf117.pep	AGINTTHVDDASKHTGRSGGNGKLVITDKAQSHHETAQSSTFEGKQVVLQAGNDANILGS	90
10	orf117ng	SGIHAGQVDDASKHTGRSGGNGKLVITDKAQSHHETAQSSTFEGKQVVLQAGNDANILGS	540
	orf117.pep	NVISDNGTQIQAGNHVRIGTTQTQSQSEYHQTKSGLMSAGIGFTIGSKTNTQENQSQS	150
	orf117ng	NVISDNGTRIAGNHVRIGTTQTQSQSEYHQTKSGLMSAGIGFTIGSKTNTQENQSQS	600
15	orf117.pep	NEHTGSTVGSGLKGDTTIVAGKHYESIGSTVSSPEGNNTIYAQSIDIAAHNKLNSNTTQT	210
	orf117ng	NEHTGSTVGSGLKGDTTIVASKHYEQTGSNVSSPEGNNLISTQSMDIGAAQNQLNSKTTQT	660
20	orf117.pep	YEQKXLTVAFSSPVTDLAQQ	230
	orf117ng	YEQKGLTVAFSSPVTDLAQQAIAVAHKAQKQFDKAKTTALMPWRLPMQVGRLEFKQAKAPK	720

An ORF117ng nucleotide sequence <SEQ ID 519> was predicted to encode a protein having amino acid sequence <SEQ ID 520>:

25	1	..LLVQTEKDGL HNEQTFGEKK VSENGKLHN YWRARRKGHD ETGHREQNYT
	51	LPEEITRDIS LGSFAYESH KALSRHAPSQ GTELPQSNRD NIRTAKSNGI
	101	SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPRFANYRQ WLGSDYMLGS
	151	LKLDPNNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
	201	NGATAARSMN LSVGIALSAE QAAQLTSDIV WLVOKEVKLP DGGTQTVLMP
30	251	QVYVRVKNGG IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
	301	DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLNAGN INNQSTAKSS
	351	QNAQGSSTYL DRMAGIYITG KEGVLAAQA GKDINIAGQ ISNQSDQGQT
	401	RLQAGRDINL DTVQTGKYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTLT
35	451	SGNNLNAAAEVGSAGKTLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG
	501	GNKLVITDKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVISDNGTRI
	551	QAGNHVRIGT TQTQSQSEY HQTKSGLMS AGIGFTIGSK TNTQENQSQS
	601	NEHTGSTVGS LKGDTTIVAS KHYEQTGSNV SSPEGNNLIS TQSMDIGAAQ
	651	NQLNSKTTQT YEQKGLTVAF SSPVTDLAQQ AIAVAHKAQ QFDKAKTTAL
	701	MPWRLPMQVG RLFQAKAPK K*

40 Further work revealed the following gonococcal partial DNA sequence <SEQ ID 521>:

	1	TTGCTTGTGC AAACAGAAAA AGACGGTTTG CATAACGAGC AAACCTTTGG
	51	CGAGAAGAAA GTCTTCAGCG AAAATGGTAA GTTGCAAC TACTGGCGTG
	101	CGCGTCGTAA AGGACATGAT GAAACAGGGC ATCGTGAACA AAATTATACT
45	151	TTGCCGGAGG AAATCACACG CGACATTTCA CTGGGTTCAT TTGCCTATGA
	201	ATCGCATAGC AAAGCATTAA GCCGTCATGC GCCAGCCAA GGCACGTAGT
	251	TGCCACAAAG TAACCGGAT AATATCCGTA CTGCGAAAAG CAACGGTATT
	301	TCGCTACCCT ATACGCCCAA TTCTTTTACC CCATTACCCG GCAGCAGCTT
	351	ATACATTATC AATCCTGCCA ATAAAGGCTA TCTTGTGAA ACCGATCCAC
	401	GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GCTGGGCAGC
50	451	CTCAAACCTAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGGTTATTA
	501	CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC
	551	GTTTAGACGG TTATCAAAAC GACGAAGAAC AATTTAAAGC CTTAATGGAT
	601	AATGGCGCGA CTGCGGCAGC TTCGATGAAT CTCAGCGTTG GCATTGCATT
	651	AAGTGCCGAG CAAGCAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC
55	701	AAAAAGAAAGT TAAACTTCTT GATGGCGGCA CACAAACCGT ATTGATGCCA
	751	CAGGTTTATG TACGCGTTAA AAATGGCGGC ATAGACGGTA AAGGTGCATT
	801	GTTGTCAGGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAACT
	851	CAGGCACGAT TGCAGGGCGC AATGCGCTTA TTATCAATAC CGATACGCTA
	901	GACAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC
60	951	ACAAGACATC AATAATATTG GCGGCATTCT TTCTGCCGAA CAGACATTAT
	1001	TGCTCAATGC GGGTAACAAC ATCAACAACC AAAGCACGGC CAAGAGCAGT
	1051	CAAAATGCAC AAGGTAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA

1101 TATCACAGGC AAAGAAAAAG GTGTTTATAGC AGCGCAGGCA GGCAAAGACA  
 1151 TCAACATCAT TGCCGGTCAA ATCAGCAATC AATCAGATCA AGGGCAAACC  
 1201 CGGCTGCAGG CAGGACGCGA CATTAACTTG GATACGGTAC AAACCGGCAA  
 1251 ATATCAAGAA ATCCATTTTG ATGCCGATAA CCATACCATC CGAGGTTCAA  
 1301 CGAACGAAGT CGGCAGCAGC ATTCAAACAA AAGGCGATGT TACCCtatTG  
 1351 TCAGGGAATA ATCTCAATGC CAAAGCTGCC GAAGTCGCGA GCGCAAAAGG  
 1401 CACACTTGCC GTGTATGCTA AAAATGACAT TACTATCAGC TCAGGCATCC  
 1451 ATGCCGGCCA AGTTGATGAT GCGTCCAAAC ATACAGGCAG AAGCGGCGGC  
 1501 GGTAAATAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACTGC  
 1551 TCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG  
 1601 ATGCCAACAT CCTTGGCAGT AATGTTATTT CCGATAATGG CACCCGGATT  
 1651 CAAGCAGGCA ATCATGTTTC CATTGGTACA ACCCAAATC AAAGCCAAAG  
 1701 CGAAACCTAT CATCAAACCC AAAAATCAGG ATTGATGAGT GCAGGTATCG  
 1751 GCTTCACTAT TGGCAGCAG ACAAACACAC AAGAAAACCA ATCCCAAAAGC  
 1801 AACGAACATA CAGGCAGTAC CGTAGGCAGC CTGAAAGCGC ATACCAACAT  
 1851 TGTTGCAAGC AAACACTACG AACAAACCGG CAGCAACGTT TCCAGCCCTG  
 1901 AGGCAACAA CCTTATCAGC ACGCAAAGTA TGGATATTGG CGCAGCACAA  
 1951 AACCAATTAA ACAGCAAAAC CACCCAAACC TACGAACAAA AAGGCTTAAC  
 2001 GGTGGCATTG AGTTCGCCCG TTACCGATTG GGCACAACAA GCGATTGCCG  
 2051 TAGCACACAA AGCAGCAAAC AAGTCGGACA AAGCAAAAAC GACCGCGTTA  
 2101 ATGCCATGGC GGCTGCCAAT GCAGGTTGGC AGGCCTATCA AACAGGCAAA  
 2151 GGCGCACAAA ACTTAG

This corresponds to the amino acid sequence <SEQ ID 522; ORF117ng-1>:

1 LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT  
 51 LPEEITRDIS LGSFAYESHS KALSRHAPSQ GTELPQSNRD NIRTAKSNGI  
 101 SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDFRFANYRQ WLGS DYMLGS  
 151 LKLDPNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD  
 201 NGATAARSMN LSVGIALSAE QAAQLTSDIV WLVOKEVKLP DGGTQTVLMP  
 251 QVYVRVKNKG IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL  
 301 DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLNAGNN INNQSTAKSS  
 351 QNAQGSSTYL DRMAGIYITG KEKGVLAQA GKDINIAGQ ISNQSDQGT  
 401 RLQAGRDINL DTVQTKGYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTL  
 451 SGNLNAKAA EVGSAKTLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG  
 501 GNKLVIDKA QSHHETAQSS TFEKGQVVLQ AGNDANILGS NVISDNTRI  
 551 QAGNHVRIGT TQTQSQSEY HQTKSGLMS AGIGFTIGSK TNTQENQSQS  
 601 NEHTGSTVGS LKGDITIVAS KHYEQTGSNV SSPEGNNLIS TQSMDIGAAQ  
 651 NQLNSKTTQT YEOKGLTVAE SSPVTDLAQQ AIAVAHKAAN KSDKAKTTAL  
 701 MPWRLEPMQVG RPIKQAKAHK T\*

ORF117ng-1 shows the same 90% identity over a 230aa overlap with ORF117. In addition, it shows homology with a secreted *N.meningitidis* protein in the database:

gi|2623258 (AF030941) putative secreted protein [Neisseria meningitidis] Length = 2273

Score = 604 bits (1541), Expect = e-172

Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)

Query: 1 LLVQTEKDGLHNEQTFGEKKVFSENGKLHNYWRARRKGHD ETGHREQNYTLPEEITRDIS 60  
 L+V T + L N++T G K + ++ G L H Y R +KG D TG+ Y E++ I  
 Sbjct: 739 LIVGTPESALDNDDELTKTI-TDKGDLHRYHRHHKKGRDSTGYRSRSPYEPAPEVS-SIR 796

Query: 61 LGSFAYESHKALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYII 120  
 +G AY+ + AP Q +++P + + NGI +T LP SSL+ I  
 Sbjct: 797 MGISAYKGY-----APQASDIPGTV---VPVVAENGIHPTFT-----LPNSSLFAI 840

Query: 121 NPANKGYLVETDFRFANYRQWLGS DYMLGSLKLDPNLHKRLGDGYEQR LINEQIAELT 180  
 P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYEQ+L+NEQIA+LT  
 Sbjct: 841 APNNKGYLIETDPAFTDYRKWLGS GYMLAALQDDPNH1HKRLGDGYEQKLVEQIAKLT 900

Query: 181 GHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWLVOKEVKLP 240  
 G+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQ A+LTSDIVWL + V LP  
 Sbjct: 901 GYRRLDGYTNDEEQFKALMDNGITIAKELQITPGIALSAEQVARLTSDIVWLENETVTL 960

Query: 241 DGGTQTVLMPQVYVRVKNKGIDGKGALLSGSNTQINVSGSLKN-SGTIAGRNALIINTDT 299  
 DG TQTVL P+VYVR + ++G+GALLSGS I SG+++N G IAGR ALI+N  
 Sbjct: 961 DGTQTVLKPVKYVRARPKDMNGQALLSGSVVDIG-SGAIENRGGLIAGREALILNAQN 1019

Query: 300 LDNIGGRIHAQKSAVTATQDINNIGGILSAEQTLNAGNNINNQSTAKSSQNAQGSSTY 359

-307-

+ N+ G + + A D I N G I A E L L L A N N I ++S +S+QN QGS  
 Sbjct: 1020 IKNLQGDLOQKNIFAAAGSDITNTGSI-GAENALLLKASNNIESRSETRSNQNEQGSVRN 1078  
 Query: 360 LDRMAGIYITGKEKGVLAQAQAGKDINIIAGQISNQSDQGGQTRLQAGRDINLDTVQTGKYQ 419  
 + R+AGIY+TG++ G + AG +I + A +++NQs+ GQT L AG DI DT + Q  
 Sbjct: 1079 IGRVAGIYLTGRQNGSVLLDAGNNIVLTASELTNQSEdGQTVLNAAGDIRSDTTGISRNO 1138  
 Query: 420 EIHFDADNHTIRGSTNEVGSSIQTKGDVTLTSGNNLNAKAAEVGSAKGTALAVYAKNDITI 479  
 FD+DN+ IR NEVGs+I+T+G+++L + ++ +AAEVGS +G L + A DI +  
 Sbjct: 1139 NTIFDSNYVIRKEQNEVGSTIRTRGNLSLNAKGDIRIRAAEVGSEQRLKLAAGRDIKV 1198  
 Query: 480 SSGIHAGQVDDASKHTGRSGGKNLVIITDKAQSHHETAQSSTFEGKQVVLQAGNDANILG 539  
 +G + +DA K+TGRSGGG K +T ++ + A S T +GK+++L +G D + G  
 Sbjct: 1199 EAGKAHTETEDALKYTGRSGGGIKQKMTIRHLKNQNGQAVSGTLDGKEIILVSGRDITVTG 1258  
 Query: 540 SNVISDNGTRIQAAGNHVRIGTTTQTSQSQSEYHQTQKSGLM-SAGIGFTIGSKTNTQENQS 598  
 SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S  
 Sbjct: 1259 SNIIADNHTILSAKNNIVLKAATRSRSAEMNKKEKSGLMGSGGIGFTAGSKKDTQTNRS 1318  
 Query: 599 QSNEHTGSTVGSGLKGDTTIVASKHYEQTGSNVSSPEGNNLISTQSMDIGAAQNQLNSKTT 658  
 ++ HT S VGSL G+T I A KHY QTGS +SSP+G+ IS+ + I AQON+ + ++  
 Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKHYTQTGSTISSPQGDVGISSGKISIDAAQNRYSQESK 1378  
 Query: 659 QTYEQKGLTVAFFSSPVT D 676  
 Q YEQKG+TVA S PV +  
 Sbjct: 1379 QVYEQKGVTVVAISVPVNV 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 63

30 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 523>:

1 ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCCTCGCCG TTGTGCGCTA  
 51 CAACATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTTCG  
 101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAWAACCAG CCATGTCCGC  
 151 GACGGCAAAC CGTCCGGCGG GTCAGTCATG ATGCCGAAAC CCCAACCAGC  
 201 GGTCAAAAAA ACGGCAAAAC CCCAAGACCC CGyCATGCGC AACCTGCAAG  
 251 AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG  
 301 TTCAAAACCG AAATCGAAAC CGCCTTGGAA GAAAGCGGCA TTATCGGCAA  
 351 CTCCGCCCAC ACCGTTTCCG AACCCCAAAC CGGACATTCC GCAACGAAAC  
 401 CTGCCGACGC GTCGGCAAAA CCTGCACCCG TTCCGCAAAAC ACCTGCAAAA  
 451 CCGCTGATTA CGCTCAAAGA ACTGTCAAAA GTCGAATTAT CCTGGTTTGA  
 501 CGTGCGCATC GACTTCATCT CCTAT...

This corresponds to the amino acid sequence <SEQ ID 524; ORF119>:

1 MIYIVLFLAV VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSXTSHVR  
 51 DGKPSGGSVMPKPOPAVKK TAKPQDPXMR NLQEQDAVYI AKQKQAKASP  
 101 FKTEIETALE ESGIIGNSAH TVSEPQTGHS ATKPADASAK PAPVPQTPAK  
 151 PLITLKELSK VELSWFDVRI DFISY...

Further work revealed the complete nucleotide sequence <SEQ ID 525>:

1 ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCCTCGCCG TTGTGCGCTA  
 51 CAACATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTTCG  
 101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAAAACCAG CCATGTCCGC  
 151 GACGGCAAAC CGTCCGGCGG GTCAGTCATG ATGCCGAAAC CCCAACCAGC  
 201 GGTCAAAAAA ACGGCAAAAC CCCAAGACCC CGCCATGCGC AACCTGCAAG  
 251 AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG  
 301 TTCAAAACCG AAATCGAAAC CGCCTTGGAA GAAAGCGGCA TTATCGGCAA  
 351 CTCCGCCCAC ACCGTTTCCG AACCCCAAAC CGGACATTCC GCACCGAAAC  
 401 CTGCCGACGC GTCGGCAAAA CCTGCACCCG TTCCGCAAAAC ACCTGCAAAA  
 451 CCGCTGATTA CGCTCAAAGA ACTGTCAAAA GTCGAATTAC CCTGGTTTGA  
 501 CGTGCGCTTC GACTTCATCT CCTATATCGC GCTGACCGAA GCCAAAGAAC  
 551 TGCACGCACT GCCGCGCCTT TCCAACCGCT GCCGCTACCA GATTGTGCGG  
 601 TGCACCATGG ACGACCATTT CCAGATTGCC GAACCCATCC CGGGCATCCG



5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

```

651 CTATCAGGCA TTTATCGTGG GTATTTCAGGC AGTCAGCCGC AACGGACTTG
701 CCTCGCAGGA AGAAGTCTCC GCATTCAACC GCCAGGTGGA CGCATTTCGA
751 CAAAGCATGG GCGGTCAGAC GCTGCACACC GACCTTGCCG CCTTTATCGA
801 AGTGGCTTCC GCACTGGACG CATTCTGCGC GCGCGTCGAC CAGACCATCG
851 CCATCCATTT GGTTCCTCCG ACCAGCATCA GCGGCGTAGA ACTGCGTTCC
901 GCCGTAACGG GCGTGGGTTT CGTTTGGAA GACGACGGCG CGTTCCACTA
951 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCAACAACG
1001 AGCCGTTTAC CAACGCCCTT TTGGACAACC AGTCCTACAA AGGCTTCAGT
1051 ATGCTGCTCG ACATCCCGCA CTCTCCGGCA GGCGAAAAAA CCTTCGACGA
1101 TTTGTTTATG GATTGGCGG TACGCTGTC CGGCCAGTTG AACCTGAATC
1151 TGGTCAACGA CAAATGGAA GAAGTTTCGA CCCAATGGCT CAAAGACGTG
1201 CGCACTTATG TATTGGCGG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA
1251 ACCGGGCGGC AAAACCGCAT TCGCCTGTT CTCCTAA

```

This corresponds to the amino acid sequence <SEQ ID 526; ORF119-1>:

15  
20  
25  
30  
35  
40  
45  
50  
55  
60

```

1 MIYIVLFLAV VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSHVR
51 DGKPSGGSSVM MPKPQPAVKK TAKPQDPAMR NLQEQDAVYI AKQKQAKASP
101 FKTEIETALE ESGIIGNSAH TVSEPQTGHS APKPADAPAK PAPVPQTPAK
151 PLITLKELESK VELPWFVRF DFISYIALTE AKELHALPRL SNRCRYQIVG
201 CTMDHDFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQVDAFA
251 QSMGGQTLHT DLAAFIEVAS ALDAFCARVD QTIATHLVSP TSISGVELRS
301 AVTGVGVFLE DDGAFTYDT SGSTMFSICS LNNEPFTNAL LDNQSYKGFS
351 MLLDIPHSPA GEKTFDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV
401 RTYVLARQSE MLKVGIPEGG KTALELRF*

```

Computer analysis of this amino acid sequence gave the following results:

#### 25 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF119 shows 93.7% identity over a 175aa overlap with an ORF (ORF119a) from strain A of *N. meningitidis*:

30  
35  
40  
45

```

          10      20      30      40      50      60
orf119.pep MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSXTSHVRDGKPSGGSSVM
orf119a     MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVRDGKPSGGPVM
          10      20      30      40      50      60

          70      80      90      100     110     120
orf119.pep MPKPQPAVKKTAKPQDPXMRNLQEQDAVYIAKQKQAKASPFKTEIETALESIGIIGNSAH
orf119a     MPKPQPAVKKTAQSQDPAMRNLQEQDAVYIAKQKQAKASPFKTEIETALESIGIIGNSAH
          70      80      90      100     110     120

          130     140     150     160     170
orf119.pep TVSEPQTGHSATKPADASAKPAPVPQTPAKPLITLKELESKVELSWDFVRIDFISY
orf119a     TVPEPQTGHSAPKPADAPAKVPVPQTPAKPLITLKELESKVELPWFVRFDFISYIALTE
          130     140     150     160     170     180

orf119a     AKELHALPRLSNRCRYQIVGCTMDHDFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS
          190     200     210     220     230     240

```

The complete length ORF119a nucleotide sequence <SEQ ID 527> is:

50  
55  
60

```

1 ATGATTTACA TCGTACTGTT CCTCGCCGCC GTCCTCGCCG TTGTGCGCTA
51 CAATATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCCG GACCAGTTCCG
101 GGCACCTCGA CAAAGATGCC CTGCTCAACA GCAAAACCAG CCATGTCCCG
151 GACGGCAAAC CGTCCGGCGG GCCAGTCATG ATGCCGAAAC CCCAACCGGC
201 GGTCAAAAAA ACGGCAAAAT CCCAAGACCC CGCCATGCGC AACCTGCAAG
251 AGCAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG
301 TTCAAAACCG AAATCGAAAC CGCCTTGGAA GAAAGCGGCA TTATCGGCAA
351 CTCCGCCAC ACCTTCCCG AACCCCAAAC CGGACATTCC GCACCAAAAC
401 CTGCCGACGC GCCGGCAAAA CCTGTTCCCG TTCCGCAAAC GCCGGCAAAA
451 CCGCTGATTA CGCTCAAAGA GCTGTCGAAG GTCGAGCTGC CCTGGTTTGA
501 CGTGCGCTTC GACTTCATCT CTTATATCGC GCTGACCGAA GCCAAGAAGC
551 TGCACGCACT GCCGCGCTT TCCAACCGCT GCCGTACCA GATTGTCGGC
601 TGCACCATGG ACGACCATT CCAGATTGCC GAACCCATCC CGGGCATCCG

```

-309-

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65

```

651 CTATCAGGCA TTTATCGTGG GTATTCAGGC AGTCAGCCGC AACGGACTTG
701 CCTCGCAGGA AGAACTCTCC GCATTCAACC GCCAGGTGGA TGCATTGCGA
751 CACAGCATGG GCGGTCAGAC GCTGCACACC GACCTTGCCG CCTTTATCGA
801 AGTGGCTTCC GCACTGGACG CATTCTGCGC GCGCGTCGAC CAGACTATCG
851 CCATCCATTT GGTTCCTCCG ACCAGCATCA GCGGCGTAGA ACTGCGTTCC
901 GCCGTAACGG GCGTGGGTTT CGTTTGGAA GACGACGGCG CGTTCCACTA
951 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCAACAACG
1001 AGCCGTTTAC CAATGCCCTT TTGGACAACC AGTCCTATAA AGGCTTCAGT
1051 ATGCTGCTCG ACATCCCGCA CTCTCCGGCA GGCGAAAAAA CCTTCGACGA
1101 TTTGTTTATG GATTTGGCGG TACGCCTGTC CGGCCAGTTG AACCTGAATC
1151 TGGTCAACGA CAAAATGGAA GAAGTTTCGA CCAATGGCT CAAAGACGTG
1201 CGCACTTATG TATTGGCTCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA
1251 ACCGGGCGGC AAAACCGCAT TGCCTGCTT CTCCTAA

```

This encodes a protein having amino acid sequence <SEQ ID 528>:

15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65

```

1 MIYIVLFLAA VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSHVR
51 DGKPSGGPVM MPKPQPAVKK TAKSQDPAMR NLQEQDAVYI AKQKQAKASP
101 FKTEIETALE ESGIIGNSAH TVPEPQTGHS APKPADAPAK PVPVPQTPAK
151 PLITLKELSK VELPWFDFVR DFISYIALTE AKELHALPRL SNRCRYQIVG
201 CTMDHDFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQVDAFA
251 HSMGGQTLHT DLAAFIEVAS ALDAFCARVD QTIAIHLVSP TSISGVELRS
301 AVTGVGFVLE DDGAFHYTDT SGSTMFSICS LNNEPFTNAL LDNQSYKGFS
351 MLLDIPHSPA GEKTFDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV
401 RTYVLARQSE MLKVGIEPGG KTA LRLES*

```

ORF119a and ORF119-1 show 98.6% identity in 428 aa overlap:

25  
30  
35  
40  
45  
50  
55  
60  
65

```

orfl19a.pep      10      20      30      40      50      60
MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALNSKTSHV RDGKPSGGPVM
|||||:|||||
orfl19-1          10      20      30      40      50      60
MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALNSKTSHV RDGKPSGGPVM

orfl19a.pep      70      80      90      100     110     120
MPKPQPAVKKTAKSQDPAMRN LQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH
|||||:|||||
orfl19-1          70      80      90      100     110     120
MPKPQPAVKKTAKSQDPAMRN LQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH

orfl19a.pep      130     140     150     160     170     180
TVPEPQTGHSAPKPADAPAKPVPVPTPAKPLITLKELSKVELPWFDFVRDFISYIALTE
|||:|||||
orfl19-1          130     140     150     160     170     180
TVPEPQTGHSAPKPADAPAKPVPVPTPAKPLITLKELSKVELPWFDFVRDFISYIALTE

orfl19a.pep      190     200     210     220     230     240
AKELHALPRLSNRCRYQIVGCTMDHDFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS
|||||:|||||
orfl19-1          190     200     210     220     230     240
AKELHALPRLSNRCRYQIVGCTMDHDFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS

orfl19a.pep      250     260     270     280     290     300
AFNRQVDAFAHSMGGQTLHTDLAA FIEVASALDAFCARVDQTIAIHLVSPTSISGVELRS
|||||:|||||
orfl19-1          250     260     270     280     290     300
AFNRQVDAFAHSMGGQTLHTDLAA FIEVASALDAFCARVDQTIAIHLVSPTSISGVELRS

orfl19a.pep      310     320     330     340     350     360
AVTGVGFVLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIPHSPA
|||||:|||||
orfl19-1          310     320     330     340     350     360
AVTGVGFVLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIPHSPA

orfl19a.pep      370     380     390     400     410     420
GEKTFDDL FMDLAVRLSGQLNLNLVNDKMEEVSTQWLKDV RTYVLARQSEMLKVGIEPGG
|||||:|||||
orfl19-1          370     380     390     400     410     420
GEKTFDDL FMDLAVRLSGQLNLNLVNDKMEEVSTQWLKDV RTYVLARQSEMLKVGIEPGG

```

```

orf119a.pep    KTALRLFSX
|||||
orf119-1       KTALRLFSX

```

### 5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF119 shows 93.1% identity over a 175aa overlap with a predicted ORF (ORF119ng) from *N.gonorrhoeae*:

```

10 orf119.pep    MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSXTSHVRDQKPSGGSV 60
    |||||:|||||
orf119ng       MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHV 60
    |||||

orf119.pep     MPKPQPAVKKTAKPDPMRNLEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH 120
    |||||
orf119ng       MPKPQPAVKKPAKQDSAMRNLEQDAVYIAKQKQAKASPFKTEIETALEEIGIIGNSAH 120
    |||||

15 orf119.pep     TVSEPQTGHSATKPADASAKPAPVPQTPAKPLITLKELSKVELSWDFVRIDFISY 175
    |||||
orf119ng       TVSEPQTGHSAPKPADAPAKPVPVPQTPAKPLITLKELSKVELPWDFVRDFISYIALTE 180
    |||||

```

The complete length ORF119ng nucleotide sequence <SEQ ID 529> is:

```

20 1 ATGATTTACA TCGTACTGTT CCTCGCGGCC GTCCTCGCCG TTGTCGCCTA
    51 CAATATGTAT CAGGAAAACC AATACCGCAA AAAAGTGC GC GACCAGTTTCG
    101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAAACCAG CCATGTCCGC
    151 GACGCGAAAC CGTCCGCGCG GCCAGTCATG ATGCCGAAAC CCCAACCGGC
    201 GGTCAAAAAA CCGGCCAAAC CCCAAGACTC CGCCATGCGC AACCTGCAAG
    25 251 AACAGGATGC CGTCTACATC GCAAGCAGA AACAGGCAA AGCCTCCCCG
    301 TTCAAACCG AAATCGAAAC CGCCTTGGA GAAATCGGCA TTATCGGCAA
    351 CTCGCCCCAC ACCGTTCCG AACCCCAAAC CGGACATTC GCACCGAAAC
    401 CTGCGGACGC GCCGGCAAAA CCGTTCCCG TTCCGCAAA GCGCGCAAAA
    451 CCGCTGATTA CGCTCAAAGA GCTGTGGA GTCGAGCTGC CCTGGTTTGA
    30 501 CGTGCCTtc gACTTCATCT CCTATATCGC GCTGACCGAA GCCAAAGAAC
    551 TGCACGCACT GCCGCGCCTT tccAACCGCT GCCGCTACCA GATTGTGCGC
    601 TGCACCATGG ACGACCATTT CCAGATTGCC GAACCCATCC CGGGCATCCG
    651 CTATCAGGCA TTTATCGTGG GTATCCAGGC AGTCAGCCGC AACGGACTTG
    701 CCTCGCAGGA AGAACTCTCC GCATTCAACC GCCAGGCGGA CGCATTCGCA
    35 751 CAAAGCATGG GCGGTGAGC GCTGCACACC GACCTTGCCG CCTTTATCGA
    801 AGTGGCTTCC GCACTGGACG CATCTGCGC GCGCGTGCAG CAGACCATCG
    851 CCATCCATTT GGTTCGCGC ACCAGCATCA GCGGCGTAGA ACTGCGTTCC
    901 GCCGTAACGG GCGTGGGTTT CGTTTGGAA GACGACGCG CGTTCCTACTA
    951 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCACAACG
    40 1001 AGCGGTTTAC CAATGCCCTT TTGGACAACC AGTCCTACAA AGGCTTCAGT
    1051 ATGCTGCTCG ACATCCCGCA CTCTCCGGCA GCGGAAAAA CCTTCGACGA
    1101 TTTGTTTATG GATTGCGCG TACGCTGTC CGGTCAGTTG AACCTGAATC
    1151 TGGTCAACGA CAAATGGAA GAAGTTTGA CCCAATGGCT CAAAGACGTA
    1201 CGCACTTATG TATTGGCGCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA
    45 1251 ACCGGCGCG AAAACCGCCC TCGCCTGTT TTCATAA

```

This encodes a protein having amino acid sequence <SEQ ID 530>:

```

50 1 MIYIVLFLAA VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSHVR
    51 DGKPSGGPVM MPKPQPAVKK PAKQDSAMR NLQEQDAVYI AKQKQAKASP
    101 FKTEIETALE EIGIIGNSAH TVSEPQTGHS APKPADAPAK PVPVPQTPAK
    151 PLITLKELSK VELPWDFVRF DFISYIALTE AKELHALPRL SNRCRYQIVG
    201 CTMDHDFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQADAFa
    251 QSMGGQTLHT DLAAFIEVAS ALDAFCARVD QTIHILVSP TSISGVELRS
    301 AVTGVGFLVE DDGAFHYTDT SGSTMFSICS LNNEPFTNAL LDNQSYKGFs
    351 MLLDIPHSPA GEKTFDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV
    55 401 RTYVLARQSE MLKVGIEEGG KTALRLFS*

```

ORF119ng and ORF119-1 show 98.4% identity over 428 aa overlap:

```

60 orf119ng       MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHV 60
    |||||
orf119-1       MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHV 60
    |||||

```

-311-

		70	80	90	100	110	120
	orf119ng	MPKPQPAVKKPAKQDSAMRN	LQEQDAVYIAKQKQAKAS	PFKTEIETALEEIGI	IGNSAH		
5	orf119-1	MPKPQPAVKKTAKPQDPAMRN	LQEQDAVYIAKQKQAKAS	PFKTEIETALEESGI	IGNSAH		
		70	80	90	100	110	120
	orf119ng	TVSEPQTGHSAPKPADAPAK	PVPVQTPAKPLITLKE	LSKVELPWDFVRFDF	ISYIALTE		
10	orf119-1	TVSEPQTGHSAPKPADAPAK	PAPVQTPAKPLITLKE	LSKVELPWDFVRFDF	ISYIALTE		
		130	140	150	160	170	180
	orf119ng	AKELHALPRLSNRCRYQIV	GCTMDDHFQIAEPIPG	IRYQAFIVGQAVSRN	GLASQEELS		
15	orf119-1	AKELHALPRLSNRCRYQIV	GCTMDDHFQIAEPIPG	IRYQAFIVGQAVSRN	GLASQEELS		
		190	200	210	220	230	240
	orf119ng	AFNRQADAFQSMGGQTL	LHTDLAAFI	EVASALDAFCARVD	QTTIAIHLV	SPTSISG	VELRS
20	orf119-1	AFNRQADAFQSMGGQTL	LHTDLAAFI	EVASALDAFCARVD	QTTIAIHLV	SPTSISG	VELRS
		250	260	270	280	290	300
	orf119ng	AVTGVG	FVLEDDGAFHYT	DTSGSTMFSICSLN	NEPFTNALLDN	QSYKGF	SMLLDIPHSPA
25	orf119-1	AVTGVG	FVLEDDGAFHYT	DTSGSTMFSICSLN	NEPFTNALLDN	QSYKGF	SMLLDIPHSPA
		310	320	330	340	350	360
	orf119ng	GEKTFDDL	FMDLAVRLSGQLN	LNLVNDKMEEV	STQWLKDV	RTYV	LARQSEMLKVGIEP
30	orf119-1	GEKTFDDL	FMDLAVRLSGQLN	LNLVNDKMEEV	STQWLKDV	RTYV	LARQSEMLKVGIEP
		370	380	390	400	410	420
	orf119ng	KTALRLFSX					
35	orf119-1	KTALRLFSX					
		429					
40	orf119ng	KTALRLFSX					
	orf119-1	KTALRLFSX					

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### 45 Example 64

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 531>

	1	..GCGCGGCACG	GCACGGAAGA	TTTCTTCATG	AACAACAGCG	ACAC.ATCAG
	51	GCAGATAGTC	GAAAGCACCA	CCGGTACGAT	GAAGCTGCTG	ATTTCCTCCA
50	101	TCGCCCTGAT	TTCATTGGTA	GTCGGCGGCA	TCGGCGTGAT	GAACATCATG
	151	CTGGTGTCCG	TTACCGAGCG	CACCAAAGAA	ATCGGCATAC	GGATGGCAAT
	201	CGGCGCGCGG	CGCGGCAATA	TTTGCAGCA	GTTTTTGATT	GAGGCGGTGT
	251	TAATCTGCGT	CATCGGCGGT	TTGGTCGGCG	TGGGTTTGTC	CGCCGCCGTC
	301	AGCCTCGTGT	TCAATCATTT	TGTAACCGAC	TTCCCGATGG	ACATTTCCGC
	351	CATGTCCGTC	ATCGGCGCGG	TCGCCTGTTT	GACCGGAATC	GGCATCGCGT
55	401	TCGGCTTTAT	GCCTGCCAAT	AAAGCAGCCA	AACTCAATCC	GATAGACGCA
	451	TTGGCACAGG	ATTGA			

This corresponds to the amino acid sequence <SEQ ID 532; ORF134>:

	1	..ARHGTEDEFFM	NNSDXIRQIV	ESTTGTMKLL	ISSIALISLV	VGGIGVMNIM
	51	LVSVTERKE	IGIRMAIGAR	RGNIXQQFLI	EAVLICVIGG	LVGVGLSAAV
60	101	SILVFNHFVTD	FPMDISAMSV	IGAVACSTGI	GIAFGFMPAN	KAAKLNPIDA
	151	LAQD*				

Further work revealed the complete nucleotide sequence <SEQ ID 533>:

```

1  ATGTCGGTGC AAGCAGTATT GGCGCACAAA ATGCGTTCGC TTCTGACGAT
51  GCTCGGCATC ATCATCGGTA TCGCGTCGGT GGTTTCCGTC GTCGCATTGG
101 GCAATGGTTC GCAGAAAAAA ATCCTTGAAG ACATCAGTTC GATAGGGACG
151 AACACCATCA GCATCTTCCC GGGGCGCGGC TTCGGCGACA GGCGCAGCGG
201 CAGGATTAAA ACCCTGACCA TAGACGACGC AAAAATCATC GCCAAACAAA
251 GCTACGTTGC TTCCGCCACG CCCATGACTT CGAGCGGCGG CACGCTGACT
301 TACCGCAACA CCGACCTGAC CGCCTCGCTT TACGGCGTGG GCGAACAATA
351 TTTTCGACGTG CGCGGACTGA AGCTGGAAAC GGGGCGGCTG TTTGACGAAA
401 ACGATGTGAA AGAAGACGCG CAGGTCGTCG TCATCGACCA AAATGTCAAA
451 GACAACTCTT TTGCGGACTC GGATCCGTTG GGTAACCACT TTTTGTTCAG
501 GAAACGCCCC TTGACCGTCA TCGGCGTGAT GAAAAAAGAC GAAAACGCTT
551 TCGGCAATTC CGACGTGCTG ATGCTTGGT CGCCCTATAC GACGGTGATG
601 CACCAATATCA CAGGCGAGAG CCACACCAAC TCCATCACCG TCAAAATCAA
651 AGACCAATGCC AATACCCAGG TTGCCGAAAA AGGGCTGACC GATCTGCTCA
701 AAGCGCGGCA CGGCACGGAA GATTTCTTCA TGAACAACAG CGACAGCATC
751 AGGCAGATAG TCGAAAGCAC CACCGGTACG ATGAAGCTGC TGATTTCCTC
801 CATCGCCCTG ATTTTCATTG TAGTCGGCGG CATCGGCGTG ATGAACATCA
851 TGCTGGTGTC CGTTACCGAG CGCACCAAAG AAATCGGCAT ACGGATGGCA
901 ATCGGCGCGC GGCGCGGCAA TATTTTGCAG CAGTTTTTGA TTGAGGCGGT
951 GTTAATCTGC GTCATCGGCG GTTTGGTCGG CGTGGGTTTG TCCGCCGCCG
1001 TCAGCCTCGT GTTCAATCAT TTTGTAACCG ACTTCCCGAT GGACATTTC
1051 GCCATGTCCG TCATCGGCGC GGTCGCCTGT TCGACCGGAA TCGGCATCGC
1101 GTTCGGCTTT ATGCCTGCCA ATAAAGCAGC CAACTCAAT CCGATAGACG
1151 CATTGGCACA GGATTGA

```

This corresponds to the amino acid sequence <SEQ ID 534; ORF134-1>:

```

1  MSVQAVLAHK MRSLLTMLGI IIGIASVSV VALNGSQKK ILEDISSIGT
51  NTISIFPGRG FGDRRSRIK TLTIDDAKII AKQSYVASAT PMTSSGGTIT
101 YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK
151 DKLFADSDPL GKTILFRKRP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM
201 HQITGESHTN SITVKIKDNA NTQVAEKGLT DLLKARHGTE DFFMNSDSI
251 RQIVESTTGT MKLLISSIAL ISLVGGIGV MNIMLVSVTE RTKEIGIRMA
301 IGARRGNILQ QFLIEAVLIC VIGGLVGVGL SAAVSLVFNH FVTDFPMDIS
351 AMSVIGAVAC STGIGIAFGE MPANKAAKLN PIDALAQD*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical protein o648 of *E. coli* (accession number AE000189)

ORF134 and o648 protein show 45% aa identity in 153aa overlap:

```

Orf134: 2  RHGTEDFFMNSDXIRQIVESTTGTMKXXXXXXXXXXXXXVVGIGVMNIMLVSVTERTKEI 61
RHG +DFF N D + + VE TT T++ VVGIGVMNIMLVSVTERT+EI
o648: 496 RHGKKDFFTWNMDGVLTVEKTTRTLQLFLTLVAVISLVVGIGVMNIMLVSVTERTREI 555

Orf134: 62  GIRMAIGARRGNIXQFLIEAXXXXXXXXXXXXXXXXXXFNHFVTDFFPMDISAMSVI 121
GIRMA+GAR ++ QFLIEA F+ + + S ++++
o648: 556 GIRMAVGARASDVLQFLIEAVLVCLVGGALGITLSLLIAFTLQLFLPGWEIGFSPLALL 615

Orf134: 122 GAVACSTGIGIAFGMPANKAAKLNPIDALAQD 154
A CST GI FG++PA AA+L+P+DALA++
o648: 616 LAFLCSTVTGILFGWLPARNAARLDPVDALARE 648

```

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF134 shows 98.7% identity over a 154aa overlap with an ORF (ORF134a) from strain A of *N. meningitidis*:

```

55  orf134.pep          10      20      30
                        ARHGTEDFFMNSDXIRQIVESTTGTMKLL
orf134a  GESHTNSITVKIKDNANTQVAEKGLTDLLKARHGTEFFMNSDSIRQIVESTTGTMKLL
                210      220      230      240      250      260

                40      50      60      70      80      90

```

-313-

```

5  orf134.pep  ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNIXQQFLIEAVLICVIGG
   orf134a    ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNILQQFLIEAVLICVIGG
                                270      280      290      300      310      320

10  orf134.pep  LVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA
   orf134a    LVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA
                                330      340      350      360      370      380

15  orf134.pep  LAQDX
   orf134a    LAQDX

```

The complete length ORF134a nucleotide sequence <SEQ ID 535> is:

```

20  1  ATGTCGGTGC AAGCAGTATT GCGGCACAAA ATGCGTTCGC TTCTGACGAT
   51  GCTCGGCATC ATCATCGGTA TCGCTTCGGT TGTCTCCGTC GTCGCATTGG
  101  GCAACGGTTC GCAGAAAAAA ATCCTTGAAG ACATCAGTTC GATAGGGACG
  151  AACACCATCA GCATCTTCCC AGGGCGCGGC TTCGGCGACA GGCGCAGCGG
  201  CAGGATTAAA ACCCTGACCA TAGACGACGC AAAAATCATC GCCAAACAAA
  251  GCTACGTTGC TTCCGCCACG CCCATGACTT CGAGCGGCGG CACGCTGACT
  301  TACCGCAATA CCGACCTGAC CGCTTCTTTG TACGGTGTGG GCGAACAATA
  351  TTTCGACGTG CGCGGGCTGA AGCTGGAAC GGGCGGCTG TTTGACGAAA
  401  ACGATGTGAA AGAAGACGCG CAGGTCGTCT TCATCGACCA AAATGTCAAA
  451  GACAAACTCT TTGCGGACTC GGATCCGTTG GGTAAAACCA TTTTGTTCAG
  501  GAAACGCCCC TTGACCGTCA TCGGCGTGAT GAAAAAAGAC GAAAACGCTT
  551  TCGGCAATTC CGACGTGCTG ATGCTTTGGT CGCCCTATAC GACGGTGATG
  601  CACCAATCA CAGGCGAGAG CCACACCAAC TCCATCACCG TCAAAATCAA
  651  AGACAATGCC AATACCCAGG TTGCCGAAAA AGGGCTGACC GATCTGCTCA
  701  AAGCGCGCGC GCGCGGCAA GATTTCTTCA TGAACAACAG CGACAGCATC
  751  AGGCAGATAG TCGAAAGCAC CACCGGTACG ATGAAGCTGC TGATTTCTCT
  801  CATCGCCCTG ATTTCATTGG TAGTCGCGCG CATCGGCGTG ATGAACATCA
  851  TGCTGGTGTC CGTTACCGAG CGCACCAAAG AAATCGGCAT ACGGATGGCA
  901  ATCGGCGCGC GCGCGGCAA TATTTTGCAG CAGTTTGTGA TTGAGGCGGT
  951  GTTAATCTGC GTCATCGGCG GTTTGGTCGG CGTGGGTTTG TCCGCCGCCG
 1001  TCAGCCTCGT GTTCAATCAT TTTGTAACCG ACTTCCCGAT GGACATTTCC
 1051  GCCATGTCCG TCATCGGCGC GGTGCGCTGT TCGACCGGAA TCGGCATCGC
 1101  GTTCGGCTTT ATGCTGCCA ATAAAGCAGC CAAACTCAAT CCGATAGATG
 1151  CATTGGCGCA GGATTGA

```

This encodes a protein having amino acid sequence <SEQ ID 536>:

```

45  1  MSVQAVLAHK MRSLLTMLGI IIGIASVVSVALNGSQKK ILEDISSIGT
   51  NTISIFPGRG FGDRRSGRIK TLTIDDAKII AKQSYVASAT PMTSSGGTLT
  101  YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK
  151  DKLFDASDPL GKTLFRKRP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM
  201  HQITGESHTN SITVKIKDNA NTQVAEKGLT DLLKARHGTE DFFMNSDSI
  251  RQIVESTTGT MKLLISSIAL ISLVVGGIGV MNIMLVSVTE RTKEIGIRMA
  301  IGARRGNILQ QFLIEAVLIC VIGGLVGVGL SAAVSLVFNH FVTDFPMDIS
  351  AMSVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD*

```

50 ORF134a and ORF134-1 show 100.0% identity in 388 aa overlap:

```

55  orf134a.pep  MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALNGSQKKILEDISSIGTNTISIFPGRG
   orf134-1     MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALNGSQKKILEDISSIGTNTISIFPGRG

  orf134a.pep  FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
   orf134-1     FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV

  orf134a.pep  RGLKLETGRLLFDENDVKEDAQVVVIDQNVKDKLFDASDPLGKTILFRKRPLTVIGVMKKD
   orf134-1     RGLKLETGRLLFDENDVKEDAQVVVIDQNVKDKLFDASDPLGKTILFRKRPLTVIGVMKKD

  orf134a.pep  ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLKARHGTE
   orf134-1     ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLKARHGTE

```

```

5  orf134a.pep  DFFMNSDSIRQIVESTTGTMKLLISSIALISLVGGIGVMNIMLVSVTERTKEIGIRMA
   orf134-1    DFFMNSDSIRQIVESTTGTMKLLISSIALISLVGGIGVMNIMLVSVTERTKEIGIRMA
   orf134a.pep  IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVAC
   orf134-1    IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVAC
10  orf134a.pep  STGIGIAFGFMPANKAAKLNPIDALAQDX
   orf134-1    STGIGIAFGFMPANKAAKLNPIDALAQDX

```

### Homology with a predicted ORF from *N.gonorrhoeae*

15 ORF134 shows 96.8% identity over a 154aa overlap with a predicted ORF (ORF134.ng) from *N. gonorrhoeae*:

```

20  orf134.pep  ARHGTEFFMNSDXIRQIVESTTGTMKLL 30
   orf134ng    GESHTNSITVKIKDNANTRVAEKGLAELLKARHGTEFFMNSDSIRQMVESTTGTMKLL 264
   orf134.pep  ISSIALISLVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNIXQQFLIEAVLICVIGG 90
   orf134ng    ISSIALISLVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNILQQFLIEAVLICIGG 324
25  orf134.pep  LVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA 150
   orf134ng    LVGVGLSAAVSLVFNHFVTDFFPMDISAASVIGAVACSTGIGIAFGFMPANKAAKLNPIDA 384
30  orf134.pep  LAQD 154
   orf134ng    LAQD 388

```

The complete length ORF134ng nucleotide sequence <SEQ ID 537> is:

```

35  1  ATGTCGGTGC AAGCAGTATT GCGGCACAAA ATGCGTTCGC TTCTGACCAT
   51  GCTCGGCATC ATCATCGGTA TCGCTTCGGT TGTCTCCGTC GTCGCGCTGG
   101 GCAACGGTTC GCAGAAAAAA ATCCTCGAAG ACATCAGTTC GATGGGGACG
   151 AACACCATCA GCATCTTCCC CGGGCGCGGC TTCGGCGACA GGCGCAGCGG
   201 CAAAATCAAA ACCCTGACCA TAGACGACGC AAAAATCATC GCCAAACAAA
   251 GCTACGTTGC CTCCGCCACG CCCATGACTT CGAGCGGCGG CACGCTGACC
   301 TACCGCAATA CCGACCTGAC CGCTTCTTTG TACGGTGTGG GCGAACAATA
   351 TTTCGACGTG CGCGGGCTGA AGCTGGAAC GGGGCGGCTG TTTGATGAGA
   40  401 ACGATGTGAA AGAAGACGCG CAAGTCGTCG TCATCGACCA AAATGTCAAA
   451 GACAAACTCT TTGCGGACTC GGATCCGTTG GGTAACCA TTTTGTTCAG
   501 GAAACGCCCC TTGACCGTCA TCGGCGTGAT GAAAAAGAC GAAAACGCTT
   551 TCGGCAATTC CGACGTGCTG ATGCTTTGGT CGCCCTATAC GACGGTGATG
   601 CACCAAATCA CAGGCGAGAG CCACACCAAC TCCATCACCG TCAAAATCAA
   651 AGACAATGCC AATACCCGGG TTGCCGAAAA AGGGCTGGCC GAGCTGCTCA
   701 AAGCACGGCA CGGCACGGAA GACTTCTTTA TGAACAACAG CGACAGCATC
   751 AGGCAGATGG TCGAAAGCAC CACCGGTACG ATGAAGCTGC TGATTTCTCTC
   801 CATCGCCCTG ATTTCAATTG TAGTCGGCGG CATCGGTGTG ATGAACATTA
   851 TGCTGGTGTC CGTTACCGAG CGCACCAAAG AAATCGGCAT ACGGATGGCA
   901 ATCGGCGCGC GCGCGGCAA TATTTTGCAG CAGTTTTTGA TTGAGGCGGT
   951 GTTAATCTGC ATCATCGGAG GCTTGGTGCG CGTAGGTTTG TCCGCCGCCG
1001 TCAGCTCGT GTTCAATCAT TTTGTAACCG ATTTCCCGAT GGACATTTCG
1051 GCGGCATCCG TTATCGGGGC GGTGCGCTGT TCGACCGGAA TCGGCATCGC
1101 GTTCGGCTTT ATGCCTGCCA ATAAGGCAGC CAAACTCAAT CCGATAGATG
1151 CATTGGCGCA GGATTGA

```

This encodes a protein having amino acid sequence <SEQ ID 538>:

```

60  1  MSVQAVLAHK MRSLLTMLGI IIGIASVSV VALNGSQKK ILEDISSMGT
   51  NTISIFFGRG FGDRRSQKIK TLTIDDAKII AKQSYVASAT PMTSSGGTILT
   101 YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK
   151 DKLFADSDPL GKTILFRKRP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM
   201 HQITGESHTN SITVKIKDNA NTRVAEKGLA ELLKARHGTE DFFMNSDSI
   251 RQMVESTTGT MKLLISSIAL ISLVGGIGV MNIMLVSVTE RTKEIGIRMA
   301 IGARRGNILQ QFLIEAVLIC IIGGLVGVGL SAAVSLVFNH FVTDFPMDIS

```

351 AASVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD\*

ORF134ng and ORF134-1 show 97.9% identity in 388 aa overlap:

```

5  orf134ng      MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALGNGSQKKILEDISSMGNTNTISIFPGRG
   orf134-1     MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALGNGSQKKILEDISSIGTNTISIFPGRG
   orf134ng      FGDRRSGKIKTLTIDDAKIIAQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
10  orf134-1     FGDRRSGRIKTLTIDDAKIIAQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
   orf134ng      RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD
   orf134-1     RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD
15  orf134ng      ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTRVAEKGLELLKARHGTE
   orf134-1     ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLELLKARHGTE
   orf134ng      DFFMNNSDSIRQVESTTGTMKLLISSIALISLVGGIGVMNIMLVSVTERTKEIGIRMA
20  orf134-1     DFFMNNSDSIRQVESTTGTMKLLISSIALISLVGGIGVMNIMLVSVTERTKEIGIRMA
   orf134ng      IGARRGNILQQFLIEAVLICIGGLVGVGLSAAVSLVFNHFVTDFFPMDISAASVIGAVAC
25  orf134-1     IGARRGNILQQFLIEAVLICIGGLVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVAC
   orf134ng      STGIGIAFGFMPANKAAKLNPIDALAQDX
   orf134-1     STGIGIAFGFMPANKAAKLNPIDALAQDX

```

30 ORF134ng also shows homology to an *E.coli* ABC transporter:

```

sp|P75831|YBJZ ECOLI HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YBJZ >gi5
(AE000189) o648; similar to YBBA_HAEIN SW: P45247 [Escherichia coli] Length =
648
35  Score = 297 bits (753), Expect = 6e-80
   Identities = 162/389 (41%), Positives = 230/389 (58%), Gaps = 1/389 (0%)

Query: 1  MSVQAVLAHKMRSLLTMLXXXXXXXXXXXXXXXXXGNGSQKKILEDISSMGNTNTISIFPGRG 60
      M+ +A+ A+KMR+LLTML +G+ +++ +L DI S+GTNTI ++PG+
Sbjct: 260 MAWRALAANKMRTLTLTMLGIIIGIASVVSIVVGDAAKQMVLAIRSIGTNTIDVYPGKD 319

40  Query: 61 FGDRRSGKIKTLTIDDAKIIAQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV 120
      FGD + L DD I KQ +VASATP S L Y N D+ AS GV YF+V
Sbjct: 320 FGDDDPQYQQALKYDDLIAIQKQPWVASATPAVSQNLRLRYNNVDVAASANGVSGDYFNV 379

45  Query: 121 RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFAD-SDPLGKTILFRKRPLTVIGVMKK 179
      G+ G F++ + AQVVV+D N + +LF +D +G+ IL P VIGV ++
Sbjct: 380 YGMTFSEGNTFNQEQLNGRAQVVVLDSENTRQLFPHKADVGEVILVGNMPARVIGVAEE 439

50  Query: 180 DENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTRVAEKGLELLKARHGT 239
      ++ FG+S VL +W PY+T+ ++ G+S NSITV++K+ ++ AE+ L LL RHG
Sbjct: 440 KQSMFGSSKVLRVWLPYSTMSGVRMGQSWLNSITVRVKEGFDSEAEQQQLTRLLSLRHGK 499

55  Query: 240 EDFFMNNSDSIRQVESTTGTMKXXXXXXXXXXXXXVGGIGVMNIMLVSVTERTKEIGIRM 299
      +DFF N D + + VE TT T++ VGGIGVMNIMLVSVTERT+EIGIRM
Sbjct: 500 KDFFTWNMDGVLTVEKTTRTLQLFLTLVAVISLVGGIGVMNIMLVSVTERTREIGIRM 559

60  Query: 300 AIGARRGNILQQFLIEXXXXXXXXXXXXXXXXXXXXXFNHFVTDFFPMDISAASVIGAVA 359
      A+GAR ++LQQFLIE F+ + + S +++ A
Sbjct: 560 AVGARASDVLQQFLIEAVLVCLVGGALGITLSLLIAFTLQLFLPGWEIGFSPLALLLAFL 619

Query: 360 CSTGIGIAFGFMPANKAAKLNPIDALAQD 388
      CST GI FG++PA AA+L+P+DALA++
Sbjct: 620 CSTVTGILFGWLPARNAARLDPVDALARE 648

```



Based on this analysis, including the presence of the leader peptide and transmembrane regions in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 65

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 539>:

```

1  ..GGGACGGGAG CGATGCTGCT GCTGTTTAC GCGGTAACGA T.CTGCCCTT
51  GGCCACTGGC GTTACCCTGA GTTACACCTC GTCGATTTT TTGGCGGTAT
101 TTTCCCTCCT GATTTTGAAA GAACGGATT CCGTTTACAC GCAGGCGGTG
151 CTGCTCCTTG GTTTTGCCGG CGTGGTATTG CTGCTTAATC CCTCGTTCCG
10  201 CAGCGGTGAG GAAACGGCGG CACTCGCCGG GCTGGCGGGC GGCAGCATGT
251 CCGGCTGGGC GTATTTGAAA GTGCGGAAC TGTCTTTGGC GGGCGAACCC
301 GGCTGGCGCG TCGTGTTTTA CCTTCCGTG ACAGGTGTGG CGATGTCGTC
351 GGTTTGGGCG ACGCTGACCG GCTGGCACAC CCTGTCTTT CCATCGGCAG
401 TTTATCTGTC GTGCATCGGC GTGTCGCGC TGATTGCCCA ACTGTCGATG
15  451 ACGCGCGCCT ACAAAGTCGG CGACAAATTC ACGGTTGCCT CGCTTCTCTA
501 TATGACCGTC GTTTTTCCG CTCTGTCTGC CGCATTTTTT CTGGGCGAAG
551 AGCTTTTCTG GCAGGAAATA CTCGGTATGT GCATCATCAT CCTCAGCGGT
601 ATTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 540; ORF135>:

```

20  1  ..GTGAMLLLFY AVTILPLATG VTLSTSSIF LAVFSFLILK ERISVYQAV
51  LLLGFAGVVL LLNPSFRSGQ ETAALAGLAG GAMSGWAYLK VRELSLAGEP
101 GWRVVFYLSV TGVAMSSVWA TLTGWHTLSF PSAVYLSCIG VSALIAQLSM
151 TRAYKVGDKF TVASLSYMTV VFSALSAAFF LGEELFWQEI LGMCIIISAV
201 F*

```

- 25 Further work revealed the complete nucleotide sequence <SEQ ID 541>:

```

1  ATGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGGC
51  GGCGGCTGCG TTACCATTA TGAACGTATT GATTAAAGAG GCATCGGCAA
101 AATTTGCCCT CGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTTCA
151 ACCGTTGCGC TCGGGGCTGC CGCCGTATTG CGTCGGGACA mCTTCCGCAC
30  201 GCCCATTTGG AAAAACCCT TAAACCGCAG TATGGTCGGG ACGGGGGCGA
251 TGCTGCTGCT GTTTACGCG GTAACGCATC TGCCTTTGGC CACTGGCGTT
301 ACCCTGAGTT ACACCTCGTC GATTTTTTTG GCGGTATTTT CCTTCCTGAT
351 TTTGAAAGAA CGGATTTCCG TTTACACGCA GGCGGTGCTG CTCCTTGTT
401 TTGCCGGCGT GGTATTGCTG CTTAATCCCT CGTTCCGCAG CGGTCAGGAA
35  451 ACGGCGGCAC TCGCGGGCT GCGGGCGGC GCGATGTCCG GCTGGGCGTA
501 TTTGAAAGTG CGGAACTGT CTTTGGCGGG CGAACCCGGC TGGCGCGTCG
551 TGTTTTACCT TTCCGTGACA GGTGTGGCGA TGTCGTGGT TTGGGCGACG
601 CTGACCGGCT GGCACACCCT GTCCTTTCCA TCGGCAGTTT ATCTGTCGTG
651 CATCGGCGTG TCGCGCTGA TTGCCCACT GTCGATGACG CGCGCTACA
40  701 AAGTCGGCGA CAAATTCACG GTTGCTCGC TTTCTATAT GACCGTCGTT
751 TTTTCCGCTC TGTCTGCCG ATTTTCTCTG GGCGAAGAGC TTTTCTGGCA
801 GGAAATACTC GGTATGTGCA TCATCATCCT CAGCGGTATT TTGAGCAGCA
851 TCCGCCCCAC TGCCTTCAA CAGCGGCTGC AATCCCTGTT CCGCAAAGA
901 TAA

```

- 45 This corresponds to the amino acid sequence <SEQ ID 542; ORF135-1>:

```

1  MDTAKKDILG SGWMLVAAAC FTIMNVLIKE ASAKFALGSG ELVFWRLFS
51  TVALGAAAVL RDXFRTPHW KNHLNRSVMG TGAMLLLFYA VTHLPLATGQ
101 TSYTSSIFL AVFSFLILKE RISVYQAVL LLGFAGVLL LNPSFRSGQE
151 TAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSVT GVAMSSVWAT
50  201 LTGWHTLSFP SAVYLSCIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV
251 FSALSAAFFL GEELFWQEI GMCIIISGI LSSIRPTAFK ORLQSLFRQR
301 *

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF135 shows 99.0% identity over a 197aa overlap with an ORF (ORF135a) from strain A of *N. meningitidis*:

```

5      orf135.pep      10      20      30
                        GTGAMLLLFYAVTILPLATGVTLSTSSIF
                        |||||
orf135a      STVALGAAVLRDTRTPHWKNLNRSMVGTGAMLLLFYAVTHLPLATGVTLSTSSIF
              50      60      70      80      90      100

10     orf135.pep      40      50      60      70      80      90
                        LAVFSFLILKERISVYTOAVLLLGAGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLK
                        |||||
orf135a      LAVFSFLILKERISVYTOAVLLLGAGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLK
              110     120     130     140     150     160

15     orf135.pep      100     110     120     130     140     150
                        VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSM
                        |||||
orf135a      VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSM
              170     180     190     200     210     220

20     orf135.pep      160     170     180     190     200
                        TRAYKVGDKETVASLSYMTVVFSALSAAFFLGEELFWQEIILGMCIIISAVFX
                        |||||
orf135a      TRAYKVGDKETVASLSYMTVVFSALSAAFFLGEELFWQEIILGMCIIISGILSSIRPTAF
              230     240     250     260     270     280

orf135a      KQRLQSLFRQRX
              290     300

```

30 The complete length ORF135a nucleotide sequence <SEQ ID 543> is:

```

1  ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGCC
51 GCGGCCTGCG TTTACCATTA TGAACGTATT GATTAAAGAG GCATCGGCAA
101 AATTTGCCCT CGGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTTCA
151 ACCGTTGCGC TCGGGGCTGC CGCCGTATTG CGTCGGGACA CCTTCCGCAC
35 201 GCCCCATTGG AAAAACCACT TAAACCGCAG TATGGTCGGG ACGGGGGCGA
251 TGCTGCTGCT GTTTACGCG GTAACGCATC TGCCTTTGGC CACCGCGGTT
301 ACCCTGAGTT ACACCTCGTC GATTTTTTGG GCGGTATTTT CCTTCCTGAT
351 TTTGAAAGAA CGGATTTCCG TTTACACGCA GCGGTGCTG CTCCTTGTTT
401 TTGCCGCGCT GGTATTGCTG CTTAATCCCT CGTTCGCGAG CGGTACAGAA
45 451 ACGGCGGCAC TCGCCGGGCT GCGGGGCGGC GCGATGTCCG GCTGGGCGTA
501 TTTGAAAGTG CGCGAACTGT CTTTGGCGGG CGAACC CGG TGGCGCGTCG
551 TGTTTTACCT TTCCGTGACA GGTGTGGCGA TGTCATCGGT TTGGGCGACG
601 CTGACCGGCT GGCACACCCT GTCCTTTCCA TCGGCAGTTT ATCTGTCTGT
651 CATCGGCGTG TCCGCGCTGA TTGCCCACT GTCGATGACG CGCGCCTACA
45 701 AAGTCGCGCA CAAATTCACG GTTGCCCTCGC TTTCTATAT GACCGCTGTT
751 TTTTCCGCTC TGTCTGCCG ATTTTCTCTG GCCGAAGAGC TTTTCTGGCA
801 GGAAATACTC GGTATGTGCA TCATCATCCT CAGCGGTATT TTGAGCAGCA
851 TCCGCCCCAC TGCCTTCAA CAGCGGCTGC AATCCCTGTT CCGCAAAGA
901 TAA

```

50 This encodes a protein having amino acid sequence <SEQ ID 544>:

```

1  MDTAKKDILG SGWMLVAAAC FTIMNVLIKE ASAKFALGSG ELVFWRMFLS
51 TVALGAAAVL RRDTRTPHW KNHLNRSMVG TGAMLLLFYA VTHLPLATGV
101 TLSYTSSIFL AVFSFLILKE RISVYTOAVL LLGFAGVLL LNPSFRSGQE
151 TAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSVT GVAMSSVWAT
55 201 LTGWHTLSFP SAVYLSGIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV
251 FSALSAAFFL AEELFWQEIL GMCIIILSGI LSSIRPTAFK QRLQSLFRQR
301 *

```

ORF135a and ORF135-1 show 99.3% identity in 300 aa overlap:

```

60 orf135a.pep      MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMFLSTVALGAAAVL
orf135-1          MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMFLSTVALGAAAVL

```

5 orf135a.pep RRDTFRTPHWKNHLNRSMVGTGAMLLLFYAVTHLPLATGVTLSYTSSIFLAVFSFLILKE  
 orf135-1 RRDXFRTPHWKNHLNRSMVGTGAMLLLFYAVTHLPLATGVTLSYTSSIFLAVFSFLILKE  
 orf135a.pep RISVYTQAVLLLGAGVLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG  
 orf135-1 RISVYTQAVLLLGAGVLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG  
 10 orf135a.pep WRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSMTRAYKVGDKFT  
 orf135-1 WRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSMTRAYKVGDKFT  
 15 orf135a.pep VASLSYMTVVFSAALAEELFWQEILGMCIIILSGILSSIRPTAFKORLOSFLRQR  
 orf135-1 VASLSYMTVVFSAALAEELFWQEILGMCIIILSGILSSIRPTAFKORLOSFLRQR

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF135 shows 97% identity over a 201aa overlap with a predicted ORF (ORF135ng) from

20 *N.gonorrhoeae*:

orf135.pep GTGAMLLLFYAVTXLPLATGVTLSYTSSIF 30  
 orf135ng STVTLGAAAVLRRDTFRTPHWKNHLNRSMVGTGAMLLLFYAVTHLPLTTGVTLSYTSSIF 335  
 25 orf135.pep LAVFSFLILKERISVYTQAVLLLGAGVLLNPSFRSGQETAALAGLAGGAMSGWAYLK 90  
 orf135ng LAVFSFLILKERISVYTQAVLLLGAGVLLNPSFRSGQEPALAGLAGGAMSGWAYLK 395  
 30 orf135.pep VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSM 150  
 orf135ng VRELSLAGEPGWRVVFYLSATGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSM 455  
 orf135.pep TRAYKVGDKFTVASLSYMTVVFSAALAEELFWQEILGMCIIISAVF 201  
 35 orf135ng TRAYKVGDKFTVASLSYMTVVFSAALAEELFWQEILGMCIIISAAF 506

An ORF135ng nucleotide sequence <SEQ ID 545> was predicted to encode a protein having amino acid sequence <SEQ ID 546>:

1 MPSEKAFRRH LRTASFQGLH LHHFHQKV GK CGIIGFGIHI FPTLLPAAQG  
 51 ILDIQLGLFR IDFAALAVYR RTQVDFIHTV IDGIASDQAF SEVVQILRRL  
 40 101 NLGHFTDTHL IAQARRFIAD FGNIRPMRRG EAKTFRCRFR FDGIDGIGHD  
 151 FRQCGHINRL APGKDCRNGK RDKVFFHTRH YNQVCLEKTN CSARKIKFRH  
 201 QKQAKTHSTS LAARFTIRPS LSQRPFMDTA KKDILGSGWM LVAAACFTVM  
 251 NVLIKEASAK FALGSGELVF WRMLFSTVTL GAAAVLRRDT FRTPHWKNHL  
 301 NRSMVGTGAM LLLFYAVTHL PLTTGVTLSY TSSIFLAVFS FLILKERISV  
 45 351 YTOAVLLLG AGVLLNPS FRSGQEPAL AGLAGGAMSG WAYLKVRELS  
 401 LAGEPGWRV FYLSATGVAM SSVWATLTGW HTLSFPSAVY LSGIGVSALI  
 451 AQLSMTRAYK VGDKFTVASL SYMTVVFSA SAAFFLGEEL FWQEILGMCII  
 501 IISAAF\*

Further work revealed the following gonococcal sequence <SEQ ID 547>:

50 1 ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGGC  
 51 GCGCGCCTGC TTCACCGTTA TGAACGTATT GATTAAAGAG GCATCGGCAA  
 101 AATTGGCCCT CGGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTCA  
 151 ACCGTTACGC TCGGTGCTGC CGCCGTATTG CGGCGCGACA CCTTCCGCAC  
 201 GCCCATTGG AAAAACCCT TAAACCGCAG TATGGTCGGG ACGGGGGCGA  
 55 251 TGCTGCTGCT GTTTTACGCG GTAACGCATC TGCCTTTGAC AACCGGCGTT  
 301 ACCCTGAGTT ACACCTCGTC GATTTTTtgg GCGGTATTTT CCTTCCTGAT  
 351 TTTGAAAGAA CGGATTTCCG TTTACACGCA GCGGCTGCTG CTCCTTGCTT  
 401 TTGCCGGCGT GGTATTGCTG CTTAATCCCT CGTTCGCGAG CGGTCAGGAA  
 451 CCGGCGGCAC TCGCGGGGCT GCGGGGCGGC GCGATGTCCG GCTGGGCGTA  
 60 501 TTTGAAAGTG CGGAACTGT CTTTGGCGGG CGAACC CGGC TGGCGGCTCG  
 551 TGTTTTACCT TTCCGCAACC GCGGTGCGCA TGTCGT Cggt ttgggcagc  
 601 Ctgaccggct ggCACAccct GTCCTTTcca tcggcagttt ATCgtCGGG

-319-

651 CATCGGCGTG tccgcgCtgA TTGCCCAaCT GtcgatgAcg cGCGcctaca  
 701 aaGTCGGCGA CAAATTCACG GTTGCCCTCGC tttcctaTat gaccgtcGTC  
 751 TTTTCCGCCG TGTCTGCCCG ATTTTCTCTg ggcgaagagc tttTctggCA  
 801 GGAAATACTC GGTATGTGCA TCATTatccT CAGCGGCATT TTGAGCAGCA  
 851 TCCGCCCCAT TGCCTTCAAA CAGCGGCTGC AAGCCCTCTT CCGCCAAAGA  
 901 TAA

This corresponds to the amino acid sequence <SEQ ID 548; ORF135ng-1>:

1 MDTAKKDILG SGWMLVAAAC FTMNVLIKE ASAKFALGSG ELVFWRM LFS  
 51 TVTLGAAAVL RRDTRTPHW KNHLNRSVMG TGAMLLLFYA VTHLPLTTGV  
 101 TLSYTSSIFL AVESFLILKE RISVYTQAVL LLGFAGVVLL LNPFSRSGQE  
 151 PAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSAT GVAMSSVWAT  
 201 LTGWHTLSFP SAVYLSGIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV  
 251 FSALSAAFFL GEELFWQEIL GMCIIILSGI LSSIRPIAFK QRLQALFRQR  
 301 \*

15 ORF135ng-1 and ORF135-1 show 97.0% identity in 300 aa overlap:

orf135ng-1.pep MDTAKKDILGSGWMLVAAACFTVMNVLIKEASAKFALGSGELVFWRM LFSVTTLGAAAVL  
 orf135-1 MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRM LFSVTALGAAAVL  
 20 orf135ng-1.pep RRDTRTPHWKNHLNRSVMGTGAMLLLFYAVTHLPLTTGVTLSTSSIFLAVFSFLILKE  
 orf135-1 RRDTRTPHWKNHLNRSVMGTGAMLLLFYAVTHLPLATGVTLSTSSIFLAVFSFLILKE  
 25 orf135ng-1.pep RISVYTQAVLLLGFGVLLLNPSFRSGQEPALAGLAGGAMSGWAYLKVRELSLAGEPG  
 orf135-1 RISVYTQAVLLLGFGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG  
 orf135ng-1.pep WRVVFYLSATGVAMSSVWATLTGWHTLSFSAVYLSGIGVSALIAQLSMT RAYKVGDKFT  
 30 orf135-1 WRVVFYLSVTGVAMSSVWATLTGWHTLSFSAVYLSGIGVSALIAQLSMT RAYKVGDKFT  
 orf135ng-1.pep VASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIILSGILSSIRPIAFKQRLQALFRQR  
 orf135-1 VASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR

35 Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 66

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 549>:

40 1 ATGAAGCGGC GTATAGCCGT CTTCGTCTCG TCCCGCAGA TAATCCGAGT  
 51 TTTGGGACAA CTGTTGCCGA AAATCGTCAA TACAGTTCCG GCACATCGGA  
 101 TGCTCTTCCA GATTTTCGGG ATGTTCTTTT TCTTCATACA CCAGCAATAT  
 151 CTGCCCCGGA TCGCCGAAAT CGATTCCCCA TCGGGCATCG TGTTCGGTGC  
 201 GCTCCTCTTC CGTCATCTGC CCGCGCATTG CCTGTATGGT AAAGCCGCCG  
 45 251 TAGGGGATGC CgTTGCACAC GAACATCCAG TCGCTGATGT CGTCAACCGG  
 301 AACGCAAACG cTTTCGCCTT GTTCGACATT GGTCAAGTTCG CCsGGTTCAT  
 351 TGTTAGCAC ACCGTAAATA TAAAGACCGT CAAAATAAAT ATCGTCGATC  
 401 CACATATGTT CGCAAATTC GCCGTCTTCG CCGTCTTGA AAAAAGGGAC  
 451 TTTGACCATG GCAAAATCCA AGGCGGAAAT AATGCGGCGG CGTTCCCAAA  
 50 501 AAAGcTCGCG CCAAAATAT TTGAATGTTT TACGGGCGCG TTCGTCGGCA  
 551 CGGTTTACCG GTTCGTCTGC CTGTTCTACA TAATAAATGA CGGAATCGCC  
 601 CATCATATCT GCTCCTCAAC GTGTACGGTA TCTGTTTGA CCTTACTGCG  
 651 GCTTCTTgcC kTCGGCATCC GATTTCGGATT TGAAAAGTTC mmrwyATTTCG  
 701 GAATAG

55 This corresponds to the amino acid sequence <SEQ ID 550; ORF136>:

1 MKRRIAVFVL FPQIIRVLGQ LLPKIVNTVP AHRMLFQIFG MFFFFIHQQY  
 51 LPGIAEIDSP CGIVFGALLF RHLPAHCLYG KAAVGDAVAH EHPVADVNR

101 NANAFALFDI GQFAXFIVQH TVNIKTVKIN IVDPHMFANF AVFAVLEKRD  
 151 FDHGKIQQGN NAAAFPKKLA PKIFECFTGA FVGTVYRFVC LFYIINDGIA  
 201 HHSAPQVRVY LFAPYCGFLP SASDSLKSS XXSE\*

Further work revealed the complete nucleotide sequence <SEQ ID 551>:

5 1 ATGATGAAGC GCGGTATAGC CGTCTTCGTC CTGTTCCCGC AGATAATCCG  
 51 AGTTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC  
 101 GGATGCTCTT CCAGATTTTC GGGATGTTCT TTTTCTTCAT ACACCAGCAA  
 151 TATCTGCCCG GGATCGCCGA AATCGATTCC CCATGCGGCA TCGTGTTCGG  
 10 201 TGCCTCCTC TTCCGTCATC TGCCCGCGCA TTGCCTGTAT GGTAAAGCCG  
 251 CCGTAGGGGA TGCCGTTGCA CACGAACATC CAGTCGCTGA TGTCTCAAC  
 301 CGGAACGCAA ACGCTTTCGC CTTGTTTCGAC ATTGGTCAGT TCGCCGGGTT  
 351 CATTGTTCAG CACACCGTAA ATATAAAGAC CGTCAAATA AATATCGTCG  
 401 ATCCACATAT GTTCGCAAT TTCGCCGTCT TCGCCGTCTT GGAAAAAGG  
 451 GACTTTGACC ATGGCAAAAT CCAAGGCGGA AATAATGCGG CGGCGTTCCC  
 15 501 AAAAAAGCTC GCGCCAAAA TATTGAATG TTTTACGGGC GCGTTCGTCG  
 551 GCACGGTTTA CCGGTTTCGTC TGCTGTCTT ACATAATAAA TGACGGAATC  
 601 GCCCATCATT CTGCTCCTCA ACGTGTACGG TATCTGTTG CACCTTACTG  
 651 CCGCTTTCTG CCTTCGGCAT CCGATTCCGA TTTGAAAAGT TCCAAATATT  
 701 CGGAATAG

20 This corresponds to the amino acid sequence <SEQ ID 552; ORF136-1>:

1 MMKRRIAEVF LFPOIIRVLG QLLPKIVNTV PAHRMLFQIF GMFFFFIHQQ  
 51 YLPGIAEIDS PCGIVFGALL FRHLPACLY GKAAGDAVA HEHPVADVNN  
 101 RNANAFALFD IGQFAGFIVQ HTVNIKTVKI NIVDPHMFAN FAVFAVLEKR  
 151 DFDHGKIQQG NAAAFPKKL APKIFECFTG AFVGTVYRFV CLFYIINDGI  
 201 AHHSAPQVRV YLFAPYCGFL PSASDSLKS SKYSE\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF136 shows 71.7% identity over a 237aa overlap with an ORF (ORF136a) from strain A of *N. meningitidis*:

30		10	20	30	40	50	59
orf136.pep		MMKRRIAEVFLFPQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGLIAEIDS					
orf136a		MMKRRIAEVFLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGLIAEIDS					
		10	20	30	40	50	60
35		60	70	80	90	100	119
orf136.pep		PCGIVFGALLFRHLPACLYGKAAVGDAVAHEHPVADVNNANAFALFDIGQFAXFIVQ					
orf136a		PCGIVFGTLLFRHXSTHCLYGKAAVGNAVAHEHPVADVNNANAFALFDIGQFAGFIVQ					
40		70	80	90	100	110	120
		120	130	140	150	160	179
orf136.pep		HTVNIKTVKINIVDPHMFANFAVFAVLEKRDHDHGKIQQGNNAAFPKKLAPKIFECFTG					
45	orf136a	HAINVKTVKINIVDPHMFANFAVFAVLEKRALTMASKXXXMRRRSQKSSRQKYLNVLRA					
		130	140	150	160	170	180
		180	190	200	210	220	230
50	orf136.pep	AFVGTVYRFVCLFYIINDGIAHH---SAPQVRVYLFAPYCGFLPSASDSLKSSXXSEX					
orf136a		R---SPARFTGLSACSTXXMTESPIISAPQVRVYLFAPYCGFLPSASDSLKSSKYSEX					
		190	200	210	220	230	

The complete length ORF136a nucleotide sequence <SEQ ID 553> is:

55 1 ATGATGAAGC GCGGTATAGC CGTCTTCGTC CTGCTCATGC AGAAAATCCG  
 51 GATTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC  
 101 GGATGCTCTT CCAGATNTTC GGGATGTTCT TTTTCTTCAT ACACCAGCAA  
 151 TACCTGCCCG GGATCGCCGA AATCGATTCC CCATGCGGCA TCGTGTTCGG  
 201 TACGCTCCTC TTCCGTCATC NGTCCACGCA TTGCCTGTAT GGTAAAGCCG  
 251 CCGTAGGGAA TGCCGTTGCA CACGAACATC CAGTCGCTGA TGTCTCAAC

301 CGGAACGCAA ACGCTTTCGC CTTGTTTCGAC ATTGGTCAGT TCGCCGGGTT  
 351 CATTGTTTCAG CACGCCATAA ATGTAAAGAC CGTCAAAATA AATATCGTCG  
 401 ATCCACATAT GTTCGCAAAAT TTCGCCNTCT TCGCCGTCTT GGAAAAAAGG  
 451 GCTTTGACCA TGGCAAAATC TAAGGNGNNA NNGATGCGGC GCGTTCCTCA  
 501 AAAAAGCTCG CGCCAAAAAT ATTTGAATGT TTTGCGGGCG CGTTCGCGCG  
 551 CACGGTTTAC CGGTTTGTCT GCCTGTCTA CATAATAAT GACGGAATCG  
 601 CCCATCATAT CTGCTCCTCA ACGTGTACGG TATCTGTTTG CACCTTACTG  
 651 CCGCTTTCTG CCTTCGGCAT CCGATTGCGA TTTGAAAAGT TCCAAATATT  
 701 CGGAATAG

10 This encodes a protein having amino acid sequence <SEQ ID 554>:

1 MMKRRIAVFV LLMQKIRILG QLLPKIVNTV PAHRMLFQXF GMFFFFIHQQ  
 51 YLPGIAEIDS PCGIVFGTLL FRHXSTHCLY GKAAVGNVA HEHPVADVNV  
 101 RNANAFALFD IGQFAGFIVQ HAINVKT VKI NIVDPHMFAN FAXFAVLEKR  
 151 ALTMASKXXX XMRRRSQKSS RQKYLNV LRA RSPARFTGLS ACST\*\*MTES  
 201 PIISAPQVRV YLFAPYCGFL PSASDSLKS SKYSE\*

ORF136a and ORF136-1 show 73.1% identity in 238 aa overlap:

		10	20	30	40	50	60
20	orf136a.pep	MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQXFGMFFFFIHQQYLP	GIAEIDS				
	orf136-1	MMKRRIAVFVLFQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLP	GIAEIDS				
		10	20	30	40	50	60
25	orf136a.pep	PCGIVFGTLLFRHXSTHCLY	80	90	100	110	120
	orf136-1	PCGIVFGALLFRHLPAHCLY	80	90	100	110	120
		70	80	90	100	110	120
30	orf136a.pep	HAINVKT VKI NIVDPHMFANFAXFAVLEKR	130	140	150	160	170
	orf136-1	HTVNIKT VKI NIVDPHMFANFAVFAVLEKR	130	140	150	160	170
		130	140	150	160	170	180
35	orf136a.pep	R---SPARFTGLSACSTXXMTESPIISAPQVRV	190	200	210	220	230
	orf136-1	AFVGTVYRFVCLFYIINDGIAHH---SAPQVRV	190	200	210	220	230
		190	200	210	220	230	

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF136 shows 92.3% identity over a 234aa overlap with a predicted ORF (ORF136ng) from *N.gonorrhoeae*:

45	orf136.pep	MKRRIAVFVLFQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLP	59
	orf136ng	MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQRYLP	60
50	orf136.pep	PCGIVFGALLFRHLPAHCLY	119
	orf136ng	PGGIVFGTLLFRHLSAHCLY	120
55	orf136.pep	HTVNIKT VKI NIVDPHMFANFAVFAVLEKR	179
	orf136ng	HTVNIKT VKI NIVDPHMFANFAVFAVLEKR	180
	orf136.pep	AFVGTVYRFVCLFYIINDGIAHHSAPQVRV	234
	orf136ng	AFAGTVYRFVCLFYIINDGIAHHTAPQVRV	235

The complete length ORF136ng nucleotide sequence <SEQ ID 555> is:

60 1 ATGATGAAGC GCGGTATAGC CGTCTTCGTC CTGCTCATGC AGAAAAATCCG  
 51 GATTTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70

```

101 G G A T G C T C T T   C C A A A T T T T C   G G G A T G T T C T   T T T T C T T C A T   A C A C C G G C A A
151 T A C C T G C C C G   G G A T C G C C G A   A A T C G A T T C C   C C A G G C G G T A   T C G T G T T C G G
201 T A C G C T C C T C   T T C C G T C A T C   T G T C C G C G C A   T T G C C T G T A C   G G T A A A G C C G
251 C C G T A G G G G A   T G C C G T T G C A   C A C G A A C A T C   C A G T C G C T G A   T G T C G C C A A C
301 C G G A A C G C A A   A C G C T T T C G C   C T T G T T C G A C   A T T G G T C A G T   C C G C C G G G T T
351 C A T T G T T C A G   C A C A C C G T A A   A T A T A A A G A C   C G T C A A A A T A   A A T A T C G T C G
401 A T C C A C A T A T   G T T C G C A A A T   T T C G C C G T C T   T C G C C G T C T T   G G A A A A A A G G
451 G A C T T T G A C C   A T G G C A A A A T   C C A A G G C G G A   A A T A A T G C G G   C G G C G T T C C C
501 A A A A A A G C T C   G C G C C A A A A G   T A T T T G A A T G   T T T T A C G G G C   G C G T T C G C C G
551 G C A C G G T T T A   C C G G T T C G T C   T G C C T G T T C T   A C A T A A T A A A   T G A C G G A A T C
601 G C C C A T C A T A   C T G C T C C T C A   A C G T G T A C G G   T A T C T G T T T G   C A C C T T A C C G
651 C G G T T T T C T A   C C T C C G G C A T   C C G A T T C G G A   T T T G A A A A G T   T C C A A A T A T T
701 C G G A A T A G

```

This encodes a protein having amino acid sequence <SEQ ID 556>:

15  
20  
25  
30  
35

```

1  M M K R R I A V F V   L L M Q K I R I L G   Q L L P K I V N T V   P A H R M L F Q I F   G M F F F F I H R Q
51  Y L P G I A E I D S   P G G I V F G T L L   F R H L S A H C L Y   G K A A V G D A V A   H E H P V A D V A N
101 R N A N A F A L F D   I G Q S A G F I V Q   H T V N I K T V K I   N I V D P H M F A N   F A V F A V L E K R
151 F D F H G K I Q G G   N N A A F P K K L   A P K V F E C F T G   A F A G T V Y R F V   C L F Y I I N D G I
201 A H T A P Q R V R   Y L F A P Y R G F L   P P A S D S L K S   S K Y S E *

```

20 ORF136ng and ORF136-1 show 93.6% identity in 235 aa overlap:

25  
30  
35

```

orf136ng      M M K R R I A V F V L L M Q K I R I L G Q L L P K I V N T V P A H R M L F Q I F G M F F F F I H R Q Y L P G I A E I D S
orf136-1      M M K R R I A V F V L F P Q I I R V L G Q L L P K I V N T V P A H R M L F Q I F G M F F F F I H Q Q Y L P G I A E I D S

orf136ng      P G G I V F G T L L F R H L S A H C L Y G K A A V G D A V A H E H P V A D V A N R N A N A F A L F D I G Q S A G F I V Q
orf136-1      P C G I V F G A L L F R H L P A H C L Y G K A A V G D A V A H E H P V A D V V N R N A N A F A L F D I G Q F A G F I V Q

orf136ng      H T V N I K T V K I N I V D P H M F A N F A V F A V L E K R D F D H G K I Q G G N N A A F P K K L A P K V F E C F T G
orf136-1      H T V N I K T V K I N I V D P H M F A N F A V F A V L E K R D F D H G K I Q G G N N A A F P K K L A P K I F E C F T G

orf136ng      A F A G T V Y R F V C L F Y I I N D G I A H T A P Q R V R Y L F A P Y R G F L P P A S D S L K S S K Y S E X
orf136-1      A F V G T V Y R F V C L F Y I I N D G I A H S A P Q R V R Y L F A P Y C G F L P S A S D S L K S S K Y S E X

```

Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 67

40 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 557>:

45  
50

```

1  A T G G A A A A T A   T G G T A A C G T T   T T C A A A A A T C   A G A C C G C T T T   T G G C A A T C G C
51  C G C G C C G C G   T T G C T T G C C G   C C . T G C G G A C   G G C G G G A A A T   A A T G C T G T C C
101 G C A A G C C G G T   G C A A A C C G C C   A A A C C G C C G   C A G T G G T C G G   T T T G G C A C T C
151 G G T G G C G G C G   C A T C T A A A G G   A T T G C C C A T   G T A G G T A T T A   T T A A G G T T T T
201 G A A A G A A A A C   G G T A T T C C T G   T G A A G G T G G T   T A C C G G C A C C   T C C G C A G G T T
251 C G A T T G T C G G   C A A C C T T T T   G C A T C G G G T A   T G T C G C C C G A   C C G C C T C G A C
301 T T G G A A G C C G   A A A T T T T A G G   C A A A A C C G A T   T T G G T C G A T T   T A A C C T T G T C
351 C A C C A A T G G G   T T T A T C A A A G   G C G C A A A G C T   G C A A A A T T A C   A T C A A C C G A A
401 A A C T C C G C G G   C A T G C A G A T T   C A G C A G T T T C   C C A T C A A A T T   T G C C G C C . .

```

50 This corresponds to the amino acid sequence <SEQ ID 558; ORF137>:

```

1  M E N M V T F S K I   R P L L A I A A A A   L L A A X R T A G N   N A V R K P V Q T A   K P A A V V G L A L
51  G G G A S K G F A H   V G I I K V L K E N   G I P V K V V T G T   S A G S I V G N L F   A S G M S P D R L E
101 L E A E I L G K T D   L V D L T L S T N G   F I K G A K L Q N Y   I N R K L R G M Q I   Q Q F P I K F A A . .

```

Further work revealed the complete nucleotide sequence <SEQ ID 559>:

-323-

1 ATGGAAAATA TGGAACGTT TTCAAAAATC AGACCGCTTT TGGCAATCGC  
 51 CGCCGCCGCG TTGCTTGCCG CCTGCGGCAC GGCGGGAAAT AATGCTGTCC  
 101 GCAAGCCGGT GCAAACCGCC AAACCCGCGC CAGTGGTCGG TTTGGCACTC  
 151 GGTGGCGCGC CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT  
 5 201 GAAAGAAAAC GGTATTCTCTG TGAAGGTGGT TACCGGCACA TCGGCAGGTT  
 251 CGATTGTCGG CAGCCTTTTT GCATCGGGTA TGTCGCCCCA CCGCCTCGAA  
 301 TTGGAAGCCG AAATTTTAGG CAAAACCGAT TTGGTCGATT TAACCTTGTC  
 351 CACCAAGTGT TTTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA  
 401 AAGTCGGCGG CAGGCAGATT CAGCAGTTTC CCATCAAATT TGCCGCCGTT  
 10 451 GCTACTGATT TTGAAACCGG CAAGGCCGTC GCTTTCAATC AGGGGAATGC  
 501 CGGGCAGGCT GTGCGCGCTT CCGCCGCCAT TCCCAATGTG TTCCAACCCG  
 551 TTATCATCGG CAGGCATACA TATGTTGACG GCGGTCTGTC GCAGCCCGTG  
 601 CCGTCAGTG CCGCCGCGC GCAGGGGGCG AATTTCGTGA TTGCCGTCGA  
 651 TATTTCCGCC CGTCCGGGCA AAAACATCAG CCAAGGTTTC TTCTCTTATC  
 15 701 TCGATCAGAC GCTGAACGTA ATGAGCGTTT CTGCGTTGCA AAATGAGTTG  
 751 GGGCAGGCGG ATGTGGTTAT CAAACCGCAG GTTTTGGATT TGGGTGCAGT  
 801 CGGCGGATTC GATCAGAAAA AACCGGCCAT CCGGTTGGGT GAGGAGGCAG  
 851 CACGTGCCGC ATTGCCTGAA ATCAAACGCA AACTGGCGGC ATACCGTTAT  
 901 TGA

20 This corresponds to the amino acid sequence <SEQ ID 560; ORF137-1>:

1 MENMVTFSKI RPLLAIAAAA LLAACGTAGN NAVRKPVQTA KPAAVVGALAL  
 51 GGGASKGFAH VGIIKVLKEN GIPVKVVTGT SAGSIVGSLF ASGMSPDRLE  
 101 LEAEILGKTD LVDLTLSSTG FIKGEKLQNY INRKVGGRQI QQFPIKFAAV  
 151 ATDFETGKAV AFNQGNAGQA VRASAAIPNV FQPVIIGRHT YVDGGLSQPV  
 201 PVSAARRQGA NFVIAVDISA RPGKNISQGF FSYLDQTLNV MSVSALQNEL  
 251 GQADVVIKPO VLDLGAUVGF DQKKRAIRLG EEAARAALPE IKRKLAARYR  
 301 \*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

30 ORF137 shows 93.3% identity over a 149aa overlap with an ORF (ORF137a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orf137.pep		MENMVTFSKIRPLLAIAAAAALLAAXRTAGNNAVRKPVQTA	KPAAVVGALALGGGASKGFAH				
35	orf137a	MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAARKPVQTA	KPAAVVGALALGGGASKGFAH				
		10	20	30	40	50	60
		70	80	90	100	110	120
40	orf137.pep	VGIIKVLKENGIPVKVVTGTSAGSIVGNLFASGMSPDRLELEAEILGKTDLVDLTLSSTNG					
	orf137a	VGIIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTLSSTSG					
		70	80	90	100	110	120
		130	140	149			
45	orf137.pep	FIKGAKLQNYINRKLGRMQIQFPIKFAA					
	orf137a	FIKGEKLQNYINRKVGRRIQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV					
		130	140	150	160	170	180

The complete length ORF137a nucleotide sequence <SEQ ID 561> is:

50 1 ATGGAAAATA TGGAACGTT TTCAAAAATC AGACCGCTTT TGGCAATCGC  
 51 CGCCGCCGCG TTGCTTGCCG CCTGCGGCAC GGCGGGAAAT AATGCTGCCC  
 101 GCAAGCCGGT GCAAACCGCC AAACCCGCGC CAGTGGTCGG TTTGGCACTC  
 151 GGTGGCGCGC CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT  
 201 GAAAGAAAAC GGTATTCTCTG TGAAGGTGGT TACCGGCACA TCGGCAGGTT  
 55 251 CGATAGTCGG CAGCCTTTTT GCATCGGGTA TGTCGCCCCA CCGCCTCGAA  
 301 TTGGAAGCCG AAATTTTAGG TAAAACCGAT TTGGTCGATT TAACCTTGTC  
 351 CACCAAGTGT TTTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA  
 401 AAGTCGGCGG CAGGCGGATT CAGCAGTTTC CCATCAAATT TGCCGCCGTT  
 451 GCTACTGATT TTGAAACCGG CAAGGCCGTC GCTTTCAATC AAGGGGAATGC  
 60 501 CGGGCAGGCT GTGCGCGCTT CCGCCGCCAT TCCCAATGTG TTCCAACCCG  
 551 TTATCATCGG CAGGCATACA TATGTTGACG GCGGTCTGTC GCAGCCCGTG



5  
601 CCCGTCAGTG CCGCCCGCG GCANGNNNG NATNTCGTGA TTGCCGTCGA  
651 TATTTCCGCC CGTCCGAGCA AAAACATCAG CCAAGGCTTC TTCTCTTATC  
701 TCGATCAGAC GCTGAACGTA ATGAGCGTTT CCGCGTTGCA AAATGAGTTG  
751 GGGCAGGCGG ATGTGGTTAT CAAACCGCAG GTTTTGGATT TGGGTGCAGT  
801 CGGCGGATTC GATCAGAAAA AACGCGCCAT CCGGTTGGGT GAGGAGGCAG  
851 CACGTGCCGC ATTGCCTGAA ATCAAACGCA AACTGGCGGC ATACCGTTAT  
901 TGA

This encodes a protein having amino acid sequence <SEQ ID 562>:

10  
1 MENMVTFSKI RPLLAIAAAA LLAACGTAGN NAARKPVQTA KPAAVVGLAL  
51 GGGASKGFAH VGIIKVLKEN GIPVKVVTGT SAGSIVGSLF ASGMS PDRLE  
101 LEAEILGKTD LVDLTLSTSG FIKGEKLQNY INRKVGRRRI QQFPIKFAAV  
151 ATDFETGKAV AFNQGNAGQA VRASAAIPNV FQPVIIGRHT YVDGGLSQPV  
201 PVSAAARRXXX XXVIAVDISA RPSKNISQGF FSYLDQTLNV MSVSALQNEL  
251 GQADVVIKPO VLDLGA VGGF DQKKRAIRLG EEAARAALPE IKRKLAAARY  
15 301 \*

ORF137a and ORF137-1 show 97.3% identity in 300 aa overlap:

20  
orf137a.pep MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAARKPVQTA KPAAVVGLALGGGASKGFAH  
orf137-1 MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAVRKPVQTA KPAAVVGLALGGGASKGFAH  
orf137a.pep VGIIKVLKENGIPVKVVTGTSAGSIVGSLFASGMS PDRLELEAEILGKTDLVDLTLSTSG  
orf137-1 VGIIKVLKENGIPVKVVTGTSAGSIVGSLFASGMS PDRLELEAEILGKTDLVDLTLSTSG  
25  
orf137a.pep FIKGEKLQNYINRKVGRRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV  
orf137-1 FIKGEKLQNYINRKVGRRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV  
30  
orf137a.pep FQPVIIGRHTYVDGGLSQPVVPSAARRXXXXXVIAVDISARPSKNISQGFFSYLDQTLNV  
orf137-1 FQPVIIGRHTYVDGGLSQPVVPSAARRQGANFVIAVDISARPGKNISQGFFSYLDQTLNV  
35  
orf137a.pep MSVSALQNELGQADVVIKPOVLDLGA VGGFDQKKRAIRLGEEAARAALPEIKRKLAAARY  
orf137-1 MSVSALQNELGQADVVIKPOVLDLGA VGGFDQKKRAIRLGEEAARAALPEIKRKLAAARY

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF137 shows 89.9% identity over a 149aa overlap with a predicted ORF (ORF137ng) from *N.gonorrhoeae*:

40  
orf137.pep MENMVTFSKIRPLLAIAAAAALLAAXRTAGNNAVRKPVQTA KPAAVVGLALGGGASKGFAH 60  
orf137ng MENMVTFSKIRSFLAIAAAAALLAACGTAGNNAARKPVQTA KPAAVVALALGGGASKGFAH 60  
45  
orf137.pep VGIIKVLKENGIPVKVVTGTSAGSIVGNLFASGMS PDRLELEAEILGKTDLVDLTLSTNG 120  
orf137ng IGIVKVLKENGIPVKVVTGTSAGSIVGSLLASGMS PDRLELEAEILGKTDLVDLTLSTSG 120  
orf137.pep FIKGAKLQNYINRKLGMQIQFPIKFAA 149  
50  
orf137ng FIKGEKLQNYINRKVGRRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV 180

The complete length ORF137ng nucleotide sequence <SEQ ID 563> is:

55  
1 ATGGAAAATA TGGTAACGTT TTCAAAAATC AGATCATTTT TGGCAATCGC  
51 CGCCGCCGCG TTGCTTGCCG CCTGCGGTAC GGCGGGAAAC AATGCCGCC  
101 GCAAGCCGGT GCAAAACCGC AAACCCGCGC CAGTGGTCGC TTTGGCACTC  
151 GGTGGCGCGC CATCTAAAGG ATTTGCCCAT ATAGGAATTG TTAAGGTTTT  
201 GAAAGAAAAC GGTATTCCTG TGAAGGTGGT TACCGGCACA TCGGCAGGTT  
251 CGATAGTCGG CAGCCTTTTG GCATCGGGTA TGTCGCCCGA CCGCCTCGAA  
301 TTGGAAGCCG AGATTTTAGG TAAACCGAT TTAGTCGATT TAACCTTGTC  
351 CACCACTGGT TTTATCAAAG GCGAAAAGCT GCAAATTAC ATCAACCGAA  
60  
401 AAGTCGGCGG CAGGCAGATT CAGCAGTTT CCATCAAATT TGCCGCCGTT  
451 GCCACTGATT TTGAAACCGG CAAGGCCGTC GCTTTCAATC AAGGGAATGC

-325-

501 CCGGCAGGCG GTTCGTGCTT CCGCCGCCAT TCCCAATGTG TTCCAGCCAG  
 551 TCATCATCGG CAGGCACAAA TATGTTGACG GCGGTCTGTC GCAGCCCGTG  
 601 CCCGTCAGTG CCGCTCGGCG GCAGGGGGCG AATTTCGTGA TTGCCGTCGA  
 651 TATTTCCGCA CGTCCGAGCA AAAATGTCGG TCAAGGTTTC TTCTCTTATC  
 701 TCGATCAGAC GCTGAACGTG ATGAGCGTTT CCGTGTGCA AAACGAGTTG  
 751 gggcAGGCGG ATGTGGTTAT CAAACCGCag gtTTTGGATT TGGGTGCAGT  
 801 CGGCGGATTC GATCAGAAAA AGCGCGCCAT CCGGTTGGGC GAGGAGGCAG  
 851 CACGTGCCGC ATTGCCTGAA ATCAAACGCA AACTGGCGGC ATACCGTTAT  
 901 TGA

10 This encodes a protein having amino acid sequence <SEQ ID 564>:

1 MENMVTFSKI RSFLAIAAAA LLAACGTAGN NAARKPVQTA KPAAVVALAL  
 51 GGGASKGFAH IGIVKVLKEN GIPVKVVTGT SAGSIVGSLI ASGMSPDRLI  
 101 LEAEILGKTD LVDLTLSTSG FIKGEKLQNY INRKVGGRQI QQFPIKFAAV  
 151 ATDFETGKAV AFNOGNAGQA VRASAAIPNV FQPVIIGRHK YVDGGLSQPV  
 15 PVSAAARRQGA NFVIAVDISA RPSKNVGQGF FSYLDQTLNV MSVSVLQNEL  
 251 GQADVVIKPO VLDLGAVGGF DQKKRAIRLG EEAARAALPE IKRKLAAARY  
 301 \*

ORF137ng and ORF137-1 show 96.0% identity in 300 aa overlap:

20 orf137ng MENMVTFSKIRSFLLAIAAAAALLAACGTAGNNAARKPVQTAKPAAVVALALGGGASKGFAH  
 orf137-1 MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAVRKPVQTAKPAAVVLALGGGASKGFAH  
 25 orf137ng IGIVKVLKENGIPVKVVTGTSAGSIVGSLI ASGMSPDRLI LEAEILGKTD LVDLTLSTSG  
 orf137-1 VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLI LEAEILGKTD LVDLTLSTSG  
 30 orf137ng FIKGEKLQNYINRKVGGRQIQQFPIKFAAVATDFETGKAVAFNOGNAGQAVRASAAIPNV  
 orf137-1 FIKGEKLQNYINRKVGGRQIQQFPIKFAAVATDFETGKAVAFNOGNAGQAVRASAAIPNV  
 35 orf137ng FQPVIIGRHKYVDGGLSQPVPVSAARRQGANFVIAVDISARPSKNVGQGF FSYLDQTLNV  
 orf137-1 FQPVIIGRHTYVDGGLSQPVPVSAARRQGANFVIAVDISARPGKNISQGF FSYLDQTLNV  
 40 orf137ng MSVSVLQNELGQADVVIKPOVLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAARY  
 orf137 MSVSALQNELGQADVVIKPOVLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAARY

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 68

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 565>:

45 1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTTGCGAA CCGCCATGCA  
 51 CATCCTGTTG ACCGCCCTGC TCAAATGCCT CTCCCTGcTG CCGCTTTCCT  
 101 GTCTGCACAC GCTGGGAAC CCGCTCGGAC ATCTGGCGTT TTACCTTTTA  
 151 AAGGAAGACC GCGCGCGCAT CGTCGCGmAT ATGCGGCAGG CGGGTTTGAA  
 201 CCCCAGCCCC AAAACGGTCA AAGCCGTTT TGCGGAAACG GCAAAAGGCG  
 251 GTTTGGAAC TGGCCCCGCG TTTTTCAGAA AACCGGAAGA CATAGAAACA  
 50 301 ATGTTCAAAG CCGTACACGG CTGGGAACAT GTGCAGCAGG CTTTGGACAA  
 351 ACACGAAGGG CTGCTATTC..

This corresponds to the amino acid sequence <SEQ ID 566; ORF138>:

1 MFRLQFRLFP PLRTAMHILL TALLKCLSLI PLSCLHTLGN RLGHAFYLL  
 51 KEDRARIVAX MRQAGLNPD P KTVKAVFAET AKGGLELAPA FFRKPEDIET  
 101 MFKAHVHGEH VQQALDKHEG LLF

Further work revealed the complete nucleotide sequence <SEQ ID 567>:

```

      1  ATGTTTCGTT  TACAATTCAG  GCTGTTTCCC  CCTTTGCGAA  CCGCCATGCA
    51  CATCCTGTTG  ACCGCCCTGC  TCAAATGCCT  CTCCTGCTG  CCGCTTTCCT
   101  GTCTGCACAC  GCTGGGAAAC  CGGCTCGGAC  ATCTGGCGTT  TTACCTTTTA
   151  AAGGAAGACC  GCGCGCGCAT  CGTCGCCAAT  ATGCGGCAGG  CGGGTTTGAA
   201  CCCCAGCCCC  AAAACGGTCA  AAGCCGTTTT  TCGGGAACG  GCAAAAGGCG
   251  GTTTGGAAct  TGCCCCCGCG  TTTTTCAGAA  AACCAGGAAG  CATAGAAACA
   301  ATGTTCAAAG  CGGTACACGG  CTGGGAACAT  GTGCAGCAGG  CTTTGGACAA
   351  ACACGAAGGG  CTGCTATTCA  TCACGCCGCA  CATCGGCAGC  TACGATTGTT
   401  GCGGACGCTA  CATCAGCCAG  CAGCTTCCGT  TCCCGCTGAC  CGCCATGTAC
   451  AAACCGCCGA  AAATCAAAGC  GATAGACAAA  ATCATGCAGG  CGGGCAGGGT
   501  TCGCGGCAAA  GGAAAAACCG  CGCCTACCAG  CATACAAGGG  GTCAAAACAA
   551  TCATCAAAGC  CCTGCGTTCG  GGCGAAGCAA  CCATCGTCCT  GCCCGACCAC
   601  GTCCCCTCCC  CTCAAGAAGG  CGGGGAAGGC  GTATGGGTGG  ATTTCTTCGG
   651  CAAACCTGCC  TATACCATGA  CGCTGGCGGC  AAAATTGGCA  CACGTCAAAG
   701  GCGTGAAAAC  CCTGTTTTTC  TGCTGCGAAC  GCCTGCCTGG  CGGACAAGGT
   751  TTCGATTGTC  ACATCCGCCC  CGTCCAAGGG  GAATTGAACG  GCGACAAAGC
   801  CCATGATGCC  GCCGTGTTCA  ACCGCAATGC  CGAATATTGG  ATACGCCGTT
   851  TTCCGACGCA  GTATCTGTTT  ATGTACAACC  GCTACAAAAT  GCCGTAA

```

20 This corresponds to the amino acid sequence <SEQ ID 568; ORF138-1>:

```

      1  MFRLOFRLFP  PLRTAMHILL  TALLKCLSL  LPLSCLHTL  GNRLGHLAF  YLL
    51  KEDRARIVAN  MRQAGLNPD  PDKTVKAV  FAETAKGG  LELELAPA  FFRKPEDIE
   101  MFKAVHWEH  VQALDKHEG  LLFITPHI  GSYDLGG  RYISQQLP  FPLTAMY
   151  KPPKIKAI  DKIMQAG  RVRGKGK  TAPTSIQ  GKIMQAG  RVRGKGK
   201  VPSPEGEG  GVVWDF  FGKPA  YTM  TLA  AKLA  HVK  GVK  TLF  F  CCER  L  PGG  QG
   251  FDLHIRPV  QGEL  NGD  KAH  DA  AV  FNR  NAE  YW  IRR  FPT  QY  LF  MYN  RY  KMP

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF138 shows 99.2% identity over a 123aa overlap with an ORF (ORF138a) from strain A of *N.*

30 *meningitidis*:

```

      10      20      30      40      50      60
orfl38.pep  MFRLOFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARIVAX
      |||
orfl38a     MFRLOFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARIVAN
      10      20      30      40      50      60

      70      80      90     100     110     120
orfl38.pep  MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGWHEHVQALDKHEG
      |||
orfl38a     MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGWHEHVQALDKHEG
      70      80      90     100     110     120

orfl38.pep  LLF
      |||
orfl38a     LLFITPHIGSYDLGGRYISQQLPFLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTSIQG
      130     140     150     160     170     180

```

The complete length ORF138a nucleotide sequence <SEQ ID 569> is:

```

      1  ATGTTTCGTT  TACAATTCAG  GCTGTTTCCC  CCTTTGCGAA  CCGCCATGCA
    51  CATCCTGTTG  ACCGCCCTGC  TCAAATGCCT  CTCCTGCTG  CCGCTTTCCT
   101  GTCTGCACAC  GCTGGGAAAC  CGGCTCGGAC  ATCTGGCGTT  TTACCTTTTA
   151  AAGGAAGACC  GCGCGCGCAT  CGTCGCCAAT  ATGCGTCAGG  CAGGCATGAA
   201  TCCCAGCCCC  AAAACGGTCA  AAGCCGTTTT  TCGGGAACG  GCAAAAGGCG
   251  GTTTGGAAct  TGCCCCCGCG  TTTTTCAGAA  AACCAGGAAG  CATAGAAACA
   301  ATGTTCAAAG  CGGTACACGG  CTGGGAACAT  GTGCAGCAGG  CTTTGGACAA
   351  ACACGAAGGG  CTGCTATTCA  TCACGCCGCA  CATCGGCAGC  TACGATTGTT
   401  GCGGACGCTA  CATCAGCCAG  CAGCTTCCGT  TCCCGCTGAC  CGCCATGTAC
   451  AAACCGCCGA  AAATCAAAGC  GATAGACAAA  ATCATGCAGG  CGGGCAGGGT
   501  TCGCGGCAAA  GGAAAAACCG  CGCCTACCAG  CATACAAGGG  GTCAAAACAA
   551  TCATCAAAGC  CCTGCGTTCG  GGCGAAGCAA  CCATCGTCCT  GCCCGACCAC

```

-327-

5  
601 GTCCCCTCCC CTCAAGAAGG CGGGGAAGGC GTATGGGTGG ATTTCTTCGG  
651 CAAACCTGCC TATACCATGA CGCTGGCGGC AAAATTGGCA CACGTCAAAG  
701 GCGTGAAAAC CCTGTTTTTC TGCTGCGAAC GCCTGCCTGG CGGACAAGGT  
751 TTCGATTGCG ACATCCGCCC CGTCCAAGGG GAATTGAACG GCGACAAAGC  
801 CCATGATGCC GCGGTGTTCA ACCGCAATGC CGAATATTGG ATACGCCGTT  
851 TTCCGACGCA GTATCTGTTT ATGTACAACC GCTACAAAAT GCCGTAA

This encodes a protein having amino acid sequence <SEQ ID 570>:

10  
1 MFRLQFRLFP PLRTAMHILL TALLKCLSL LPLSCLHTLGN RLGHAFYLL  
51 KEDRARIVAN MRQAGLNPD KTVKAVFAET AKGGLELAPA FFRKPEDTET  
101 MFKAVHGEH VQALDKHEG LLFITPHIGS YDLGGYISQ QLPFPLTAMY  
151 KPPKIKALDK IMQAGRVGRK GKTAPTSIQG VKQIIKALRS GEATIVLPDH  
201 VPSPQEGGEG VWVDFGKPA YMTLAAKLA HVKGVKTLFF CCERLPGGQG  
251 FDLHIRPVQG ELNGDKAHDA AVFNRNAEYW IRREPTQYLF MYNRYKMP\*

ORF138a and ORF138-1 show 99.7% identity over a 298aa overlap:

15 orf138a.pep MFRLQFRLFPPLRTAMHILLTALLKCLSL LPLSCLHTLGNRLGHAFYLLKEDRARIVAN  
orf138-1 MFRLQFRLFPPLRTAMHILLTALLKCLSL LPLSCLHTLGNRLGHAFYLLKEDRARIVAN  
20 orf138a.pep MRQAGMNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGEHVQALDKHEG  
orf138-1 MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGEHVQALDKHEG  
25 orf138a.pep LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKALDKIMQAGRVGRKGKTAPTSIQG  
orf138-1 LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKALDKIMQAGRVGRKGKTAPTSIQG  
30 orf138a.pep VKQIIKALRSGEATIVLPDHVPSPQEGGEGVWVDFGKPAYMTLAAKLAHVKGVKTLFF  
orf138-1 VKQIIKALRSGEATIVLPDHVPSPQEGGEGVWVDFGKPAYMTLAAKLAHVKGVKTLFF  
35 orf138a.pep CCERLPGGQGFDLHIRPVQGEELNGDKAHDAAVFNRNAEYWIRREPTQYLFMYNRYKMP  
orf138-1 CCERLPGGQGFDLHIRPVQGEELNGDKAHDAAVFNRNAEYWIRREPTQYLFMYNRYKMP

### 35 Homology with a predicted ORF from *N.gonorrhoeae*

ORF138 shows 94.3% identity over a 123aa overlap with a predicted ORF (ORF138ng) from *N.gonorrhoeae*:

40 orf138.pep MFRLQFRLFPPLRTAMHILLTALLKCLSL LPLSCLHTLGNRLGHAFYLLKEDRARIVAX 60  
orf138ng MFRLQFRLFPPLRTAMHILLTALLKCLSL LPLSCLHTLGNRLGHAFYLLKEDRARIVAN 60  
orf138.pep MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGEHVQALDKHEG 120  
orf138ng MRQAGLNPDQTVKAVFAETAKGLELAPAFFRKPEDIETMFKAVHGEHVQALDKHEG 120  
45 orf138.pep LLF 123  
orf138ng LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKALDKIMQAGRVGRKGKTAPTSIQG 180

The complete length ORF138ng nucleotide sequence <SEQ ID 571> is:

50 1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTGCGAA CCGCCATGCA  
51 CATCCTGTTG ACCGCCCTGC TCAAATGCCT CTCCTGCTG TCGCTTTCCT  
101 GTCTGCACAC GCTGGGAAC CCGCTCGGAC ATCTGGCGTT TTACCTTTTA  
151 AAGGAAGACC GCGCGCGCAT CGTCGCCAAT ATGCGGCAGG CGGGTTTGAA  
201 CCCCACACG CAGACGGTCA AAGCCGTTT TCGGGAACG GCAAAATGCG  
55 251 GTTTGGAAC TGGCCCGCG TTTTCAA AAA AACC GAAGA CATCGAAACA  
301 ATGTTCAAAG CGGTACACG CTGGGAACAC GTGCAGCAG CTTTGGACAA  
351 GGGCGAAGGG CTGCTGTTCA TCACGCCGCA CATCGGCAGC TACGATTGCG  
401 GCGGACGCTA CATCAGCCAG CAGCTTCCGT TCCACCTGAC CGCCATGTAC  
451 AAGCCGCCGA AATCAAAGC GATAGACAAA ATCATGCAGG CGGGCAGGGT  
60 501 GCGCGGCAAA GGCAAAACcg cgccaccgg catACAAGGG GTCAAACAAA  
551 tcatcaAGGC CCTGCGCGCG GCGGAGGCAA CCatcATCCT GCCGACCA

-328-

5  
 601 GTCCCTTCTC CGCAGGAagg cggCGGCGTG TGGGCGGATT TTTTCGGCAA  
 651 ACCTGCATAc acCATGACAC TGGCGGCAAA ATTGGCACAC GTCAAAGGCG  
 701 TGAAAACCTT GTTTTCTGCG TCGAACGCC TCGCCGACGG ACAAGGCTTC  
 751 GTGTTGCACA TCCGCCCGGT CCAAGGGGAA TTGAACGGCA ACAAAGCCCA  
 801 CGATGCCGCC GTGTTCAACC GCAATACCGA ATATTGGATA CGCCGTTTTT  
 851 CGACGCAGTA TCTGTTTATG TACAACCGCT ATAAAACGCC GTAA

This encodes a protein having amino acid sequence <SEQ ID 572>:

10  
 1 MFR LQFR LFP PLRTAMHILL TALLKCLSL SLSCLHTLGN RLGH LAFYLL  
 51 KEDRARIVAN MRQAGLNPD TQTVKAVFAET AKCGLELAPA FFKKPED IET  
 101 MFKAVHGWEH VQQALDKGEG LLFITPHIGS YDLGGYISQ QLPFHLTAMY  
 151 KPPKIKAIK IMQAGRVRGK GKTAPTGIQG VKQIIKALRA GEATIILPDH  
 201 VPSPQEGGGV WADFFGKPAY TMTLAAKLAH VKGVKTLFFC CERLPDGGQGF  
 251 VLHIRPVQGE LNGNKAHDAA VFNRTYEWI RRFTQYLFM YNRYKTP\*

ORF138ng and ORF138-1 show 94.3% identity over 299aa overlap:

15  
 orf138-1.pep MFR LQFR LFPPLRTAMHILLTALLKCLSL SLSCLHTLGNRLGH LAFYLLKEDRARIVAN  
 orf138ng MFR LQFR LFPPLRTAMHILLTALLKCLSL SLSCLHTLGNRLGH LAFYLLKEDRARIVAN  
 20  
 orf138-1.pep MRQAGLNPDPTQTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGWEHVQQALDKHEG  
 orf138ng MRQAGLNPDPTQTVKAVFAETAKCGLELAPAFFRKPEDIETMFKAVHGWEHVQQALDKGEG  
 orf138-1.pep LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTISIQG  
 25 orf138ng LLFITPHIGSYDLGGYISQQLPFHLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTGIQG  
 orf138-1.pep VKQIIKALRSGEATIVLPDHVPSPQEGGGVWVDFFGKPAYTMTLAAKLAHVKGVKTLEF  
 orf138ng VKQIIKALRAGEATIILPDHVPSPQEGG-GVWADFFGKPAYTMTLAAKLAHVKGVKTLEF  
 30  
 orf138-1.pep CCERLPGGQGFVLDHIRPVQGE LINGDKAHDAAVFNRNAEYWRFRFTQYLFMYNRYKMP  
 orf138ng CCERLPDGGQGFVLDHIRPVQGE LINGNKAHDAAVFNRTYEWIRFRFTQYLFMYNRYKTP

In addition, ORF138ng is homologous to htrB protein from *Pseudomonas fluorescens*:

35  
 gnl|PID|e334283 (Y14568) htrB [*Pseudomonas fluorescens*] Length = 253  
 Score = 80.8 bits (196), Expect = 9e-15  
 Identities = 49/151 (32%), Positives = 79/151 (51%), Gaps = 6/151 (3%)  
 40  
 Query: 101 MFKAVHGWEHVQQALDKGEGLLFITPHIGSYD-LGGYISQQLPFHLTAMYKPPKIKAIK 159  
 + + V G E + + + AL G + + + IT H + G + + + L Y S Q P Y + PPK + KA + D  
 Sbjct: 94 LVREVEGLEVLKEALASGKGVVGITSHLGNWEVLNHFYCSQCKPI---IFYRPPKLKAVD 150  
 Query: 160 KIMQAGRVRGKGKTAPTGIQGVKQIIKALRAGEATIILPDHVPSPQEGGGVWADFFGKPA 219  
 + + + + RV + K A + + G + + IK + R G I D P P E G + + FF A  
 45 Sbjct: 151 ELLRKQRVQLGNKVAASTKEGILSVIKEVRKGGQVGIPAD--PEPAESAGIFVFPFATQA 208  
 Query: 220 YTMTLAAKLAHVKGVKTLEFFCERLPDGGQGF 250  
 T + + F RLPDG G +  
 Sbjct: 209 LTSKFVPNMLAGGKAVGVFLHALRLPDGSGY 239

50 Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF138-1 (57kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 14A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis

(Figure 14B). These experiments confirm that ORF138-1 is a surface-exposed protein, and that it is a useful immunogen.

### Example 69

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 573>:

```

5      1  ..GCGTGGTCGG CCGGCGAATC GTGGCGTGTG TTAATGAAA GTGAAACGTG
      51  GCATGCGGTG TGGAATACTT TCGCCTTCTC GCGGCGGCGG GTGTATGCGG
     101  CAGCGGTTTT GGGTGTGGTG TATGCGGCGC CCGCGCGGCG GTGCGGCGTG
     151  ATGCGCGGGC TGATGTTTGA GCCGTTTATG GTGTCGCCGG TTTGTGTTTC
     201  GCGGCGCGTG CTGCTGCTTT ATCCGCAGTG GACGGCTTCG TTGCCGTTGC
    10  251  TGCTGGCGAT GTATGCGCTG CTGGCGTATC CGTTTGTGGC AAAAGATGTT
     301  TTATCAGCCT GGGATGCACT GCCGCCGAT TACGGCAGGG CCGCGGCGGG
     351  TTTGGGTGCA AACGGCTTTC AGACGGCATG CCGCATCAGC TTCCCCCTCT
     401  TGAAACCGGC GTTGCGCGCG GGTCTGACTT TGGCGGCGGC AACCTGCGTG
     451  GGCGAATTTG CGGCGACATT GTTCTGTGCG CGTCCGAAT GGCAGACGCT
    15  501  GACGACTTTG ATTTATGCCT ATTTGGGACG CCGGGGTGAG GATAATTACG
     551  CGCGGGCGAT GGTGCTG..

```

This corresponds to the amino acid sequence <SEQ ID 574; ORF139>:

```

      1  ..AWSAGESWRV LMESETHAV WNTLRFSAAA VYAAAVLGVV YAAPARRSAW
     51  MRGLMFXPFM VSPVCVSAGV LLLYPQWTAS LPLLLAMYAL LAYPFVAKDV
    20  101  LSAWDALPPD YGRAAAGLGA NGFQTACRIT FPLLKPALRR GLTLAAATCV
     151  GEFAATLFLS RPEWQTLTTL IYAYLGRAGE DNYARAMVL..

```

Further work revealed the complete nucleotide sequence <SEQ ID 575>:

```

      1  ATGGATGGAC GCGGTTGGGT GGTATGGGGT GCTTTTGCCC TGCTGCCTTC
     51  GGCTTTTTTG GCGGTAATGG TCGTTGCGCC TTTGTGGGCG GTGGCGGCGT
    25  101  ATGACGGTTT GCGGTGGCGC GCGGTGCTGT CGGATGCCTA TATGCTCAAA
     151  CGTTTGGCGT GGACGGTATT TCAGGCAGCG GCAACCTGTG TGCTGGTGCT
     201  GCCTTTGGGC GTGCTGTGCG CGTGGGTGCT GCGCGGCTGT GCGTTTCCGG
     251  GCGGCGCTTT GGTGCTGCGC CTGCTGATGC TGCCTTTTGT GATGCCACCG
     301  TTGGTGGCGG GCGTGGGCGT GCTGGCCCTG TTCGGGGCGG ACGGGCTGTT
    30  351  GTGGCGCGGC AGGCAGGATA CGCCGTATCT GTTGTGTGAC GGCAATGTGT
     401  TTTTCAACCT TCCTGTGTTG GTCAGGGCGG CGTATCAGGG GTTTGTGCAA
     451  GTGCCTGCGG CACGGCTTCA GACGGCACGG ACGTTGGGCG CGGGGGCGTG
     501  GCGGCGGTTT TGGGACATTG AAATGCCCGT TTTGCGCCCG TGGCTTGCCG
     551  GCGGCGTGTG CCTTGTCTTT CTGTATTGTT TTTCCGGGTT CGGGCTGGCG
    35  601  CTGCTGCTGG GCGGCAGCCG TTATGCCACG GTCGAAGTGG AAATTTACCA
     651  GTTGGTCATG TTCGAACCTG ATATGGCGGT TGCTTCGGTG CTGGTGTGGC
     701  TGGTGTGGG GGTAAACGGC GCGGCAGGGT TGCTGTATGC GTGGTTCGGC
     751  AGGCGCGCGG TTTCGGATAA GCGGTTTCC CCTGTGATGC CGTCGCCGCC
     801  GCAGTCGGTC GGGGAATATG TGCTGCTGGC GTTGTGCGCG GCGGTGTGTG
    40  851  CTGTGTGCTG CCTGTTTCCT TTGTTGGCAA TTGTTGTGAA AGCGTGGTCG
     901  CCGGCGGAAT CGTGTCTGCT GTTAATGGAA AGTGAAACGT GGCAGGCGGT
     951  GTGGAATACT TTGCGCTTCT CCGCGGCGGC GGTGTATGCG GCGGCGGTTT
    1001  TGGGTGTGGT GTATGCGGCG GCGGCGCGGC GGTGCGCGTG GATGCGCGGG
    1051  CTGATGTTTT TGCCGTTTAT GGTGTCGCCG GTTGTGTGTT CCGCGGGCGT
    45  1101  GCTGCTGCTT TATCCGCACT GGACGGCTTC GTTGCCGTTG CTGCTGGCGA
     1151  TGTATGCGCT GCTGGCGTAT CCGTTTGTGG CAAAAGATGT TTTATCAGCC
     1201  TGGGATGCAC TGCCGCCCGA TTACGGCAGG GCGGCGGCGG GTTTGGGTGC
     1251  AAACGGCTTT CAGACGGCAT GCCGCATCAC GTTCCCCCTC TTGAAACCGG
     1301  CGTTGCGGCG CGTGTCTGCT TTGGCGGCGG CAACCTGCGT GGGCGAATTT
    50  1351  GCGGCGACAT TGTTTCTGTC GCGTCCGAA TGGCAGACGC TGACGACTTT
     1401  GATTTATGCC TATTTGGGAC GCGCGGTGGA GGATAATTAC GCGCGGGCGA
     1451  TGGTGTGAC ATTGCTGTTG GCGGCGTTCG CGTGGGTAT TTTCTGCTG
     1501  TTGGACGGCG GCGAAGGCGG AAAACAGACG GAAACGTTAT AA

```

This corresponds to the amino acid sequence <SEQ ID 576; ORF139-1>:

```

    55  1  MDGRRVVVWG AFALLPSAFL AVMVVAFLWA VAAYDGLAWR AVLSDAYMLK
      51  RLAWTVFQAA ATCVLVLP LG VPVAVVLARL AFPGRALVLR LLMLPFVMP
     101  LVAGVGLAL FGADGLLWRG RQDTPYLLLY GNVFFNLPLV VRAAYQGFVQ
     151  VPAARLQ TAR TLGAGAWRRF WDIEMPVLRP WLAGGVCLVF LYCFSGFGLA

```

5

201	<u>LLGGSR</u> YAT	VEVEIYQLVM	FELDMAVASV	<u>LVVLVLG</u> VTA	<u>AAGLLYAW</u> FG
251	RRAVSDKAVS	PVMPSPQSV	GEYVLLAFAA	<u>AVLSVCCL</u> FP	<u>LLAIVVKA</u> WS
301	AGESRWILME	SETWQAVWNT	LRFSAAAVYA	<u>AAVLGVV</u> YAA	AARRSAWMRG
351	<u>LMFLPFMV</u> SP	<u>VCVSAGV</u> LL	YPQWTASLPL	<u>LLAMYLAY</u>	PFVAKDVL
401	WDALPPDYGR	AAAGLGANGF	QTACRITFPL	LKPALRRGLT	LAATCVGEF
451	AATLFLSRPE	WQTLTLLIYA	YLGRAGEDNY	ARAMVZTL	<u>LLAAAFALGI</u> FL
501	LDGGE <del>GGK</del> QT	ETL*			

Computer analysis of this amino acid sequence gave the following results:

### Homology with a predicted ORF from *N.meningitidis* (strain A)

10 ORF139 shows 94.7% identity over a 189aa overlap with an ORF (ORF139a) from strain A of *N. meningitidis*:

```

15 orf139.pep                               AWSAGESWRVLMSESTWHAVVNTLRFSA30
                                         |||||
orf139a   QSVGEYVLLAFAAAVXSVCCFLXLLAIVVKAWSAGESWRVLMSESTWQAVVNTXRFSA300
          270      280      290      300      310      320
20 orf139.pep                               40      50      60      70      80      90
                                         VYAAAVLGVVYAAPARRSAWMRGLMFXPFMVSPVCVSAGVLLLYPQWTASLPLLAMYAL
                                         |||||
orf139a   VYAAAVLGVVYAAAARRSAWMRGLMFLPFMVSPVCVSAGVLLLYPQWTASLPLLAMYAL
          330      340      350      360      370      380
25 orf139.pep                               100     110     120     130     140     150
                                         LAYPFVAKDVLSAWDALPPDYGRAAGLGANGFQTACRITFPLLKPALRRGLTLAAATCV
                                         |||||
orf139a   LAYPFVAKDVLSAXDALPPDYGRAAGLGANGFQTACRITFPLLKPALRRGLTLAAATCV
          390     400     410     420     430     440
30 orf139.pep                               160     170     180     189
                                         GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVL
                                         |||||
orf139a   GEFAATLFXSRXEWQTLTTLIYAYXGRAGXDNYARAMVLTL450LLA460AFALGX470FLLL480DGGE490GG500
          450     460     470     480     490     500

```

35 The complete length ORF139a nucleotide sequence <SEQ ID 577> is:

	1	ATGGATGGAC	GGCGTTGGGC	GGTATGGGGT	GCTTTTGCCC	TGCTGCCTTC
	51	GGCTTTTTTG	GCGCAATGG	TCGTGCGCC	TTTGTGGGCG	GTGGCGGCGT
	101	ATGACGGTTT	GGCGTGGCGC	GCGGTGCTGT	CGGATGCCTA	TATGCTCAAA
40	151	CGTTTGGCGT	GGACGGTATT	TCAGGCAGCG	GCAACCTGTG	TGCTGGTGCT
	201	GCCTTTGGGC	GTGCCTGTGC	CGTGGGTGCT	GCGCGGGCTG	CGGTTTCCGG
	251	GGCGGGCTTT	GGTGCTGCGC	CTGCTGATGC	TGCCTTTTGT	GATGCCACG
	301	TTGGTGGCGG	GCGTGGGCGT	GCTGGCTCTG	TTGCGGGCGG	ACGGCCTGTN
	351	GTGGCGCGCG	TGGCAGGATA	CGCCGTATCT	GTTGTTGTAC	GGCAATGTGT
45	401	TTTTTNACCT	TCCTGTGTTG	GTCAGGGCGG	CATATCAGGG	GTTTGTGCAA
	451	GTGCCTGCGG	CACGGCTTCA	GACGGCACNG	ACATTGGGCG	CGGGGGCGTG
	501	GCGCGGGTTT	TGGGACATTG	AAATGCCCGT	TTTGCGCCCG	TGGCTTGCCG
	551	GCGGCGTGTG	CCTGTCTCTC	CTGTATTGTT	TTTCGGGGTT	CGGGCTGGCA
	601	TTGCTGCTGG	GCGGCAGCCG	TTATGCCACG	GTCGAAGTGG	AAATTTACCA
	651	GTTGGTCATG	TTCCAACTCG	ATATGCGGGT	TGCTTCGGTG	CNTNGTGGCG
50	701	TGGTGTNGGG	GGTAACNGCG	GCGGCAGGGT	TGCTGTATGC	GTGGTTCGGC
	751	AGCGCGCGCG	TTTCGGATAA	GGCNGTTTCC	CCTGTGATGC	CGTCGCCGCC
	801	GCAGTCCGGT	GGGGAATATG	TGCTNCTGGC	GTTTGCGGCG	GCGGTGTNGT
	851	CTGTGTGCTG	CCTGTTTCNT	TTGTTGGCAA	TGTTGTGAA	AGCGTGGTCG
	901	GCCGGCGAAT	CGTGGCGTGT	GTTAATGGAA	AGTGAACGTT	GGCAGGCGGT
55	951	GTGGAATACT	NTGCGCTTCT	CGGCGCGCGC	GGTGTATGCG	GCGGCGGTTT
	1001	TGGGTGTGTT	GTATGCGGCG	GCGGCGCGCG	GGTCCGGGTG	GATGCGCGGG
	1051	CTGATGTTTT	TGCCGTTTAT	GGTGTGCGCG	GTTTGTGTTT	CGGCGGGCGT
	1101	GCTGTGCTTT	NATCCGCAGT	GGACGGCTTC	GTTTCCGCTG	CTGCTGGCGA
	1151	TGTATGCGCT	GCTGGCGTAT	CCGTTTGTGG	CAAAAGATGT	TTTATCAGCC
	1201	TGNATGCACT	TGCCGCGCGA	TTACGCGAGG	GCGGCGGCGG	GTTTGGGTGC
	1251	AAACGGCTTT	CAGACGGCAT	GCCGCATCAC	GTTCCCCCTC	TTGAAACCGG
	1301	CGTTGCGGGC	CGGTCGTACT	GCTGCGGCGG	CACCTGCGGT	GGGCGAATTT
60	1351	GCGGCAACCT	TGTTCTNTGTC	CGGTCNCGAG	TGGCAGACGC	TGACGACTTT

1	MDGRRWAVWG	AFALLPSAFL	AAMVVAPLWA	VAAYDGLAWR	AVLSDAYMLK
51	RLAWTVFQAA	ATCVLVLPGL	VPVAWVLARL	AFPGRALVLR	LLMLPFVMPT
101	LVAGVGVLAL	FGADGLXWRG	WQDTPYLLLY	GNVFFXLPVL	VRAAYQGFVQ
151	VPAARLQTAX	TLGAGAWRRF	WDIEMPVLRP	WLAGGVCLVF	LYCFSGFGLA
201	LLGGSRYAT	VEVEIYQLVM	FELDMAVASV	LVWLXVGXTA	AAGLYAWFG
251	RRASDKAVS	PVMPSPQSV	GEYVLLAFAA	AVXSVCCLEF	LLAIVVKAWS
301	AGESWRVLME	SETWQAVWNT	XRFSAAAVYA	AAVLGVVYAA	AARRSAWMRG
351	LMFLPFMVSP	VCVSAAGVLL	XPQWTASLPL	LLAMYALLAY	PFVAKDVLSA
401	XDALPPDYGR	AAAGLGANGF	QTACRITFPL	LKPALRRGLT	LAATCVGEF
451	AATLFXSRXE	WQTLTTLIYA	YXGRAGXDNY	ARAMVLTLLL	AAFALGXFLL
501	LDGGEKKRT	ETL*			

[illegible]

ORF139 shows 95.2% identity over a 189aa overlap with a predicted ORF (ORF139ng) from

orf139.pep		AWSAGESWRVLMSESETWHAVWNTLRFSAAA	30
orf139.ng	QSVGEYVLLAFSVAVLSVCCLFPLSAIVVKAWSAGESRRVLMSESETWQAVWNTLRFSAAA	:	327
orf139.pep	VYAAAVLGVVYAAPARRSAWMRGLMFXPFMVSPVCVSAGVLLLYPQWTASLP LLAMYAL		90
orf139.ng	VFAAAVLGVVYAAAARRLVWMRGLVLPFMVSPVCVSAGVLLLYPQWTASLP LLAMYAL	:	387



orf139.pep	LAYPFVAKDVLSAWDALPPDYGRAAGLGANGFQTACRITFPLLKPALRRGLTLAAATCV	150
orf139ng	LAYPFVAKDVLSAWDALPPDYGRAAGLGANGFQTACRITFPLLKPALRRGLTLAAATCV	447
5 orf139.pep	GEFAATFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVL	189
orf139ng	GEFAATFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVLTLLLSAFAVCIFLLLDNDEGG	507

The complete length ORF139ng nucleotide sequence <SEQ ID 579> is predicted to encode a protein having amino acid sequence <SEQ ID 580>:

10	1	MDGRCWAVRG	AFSLLPSAFL	AVMVVAPLWA	VAAYDGLAWR	AVLSDAYMLK
	51	RLAWTVFQAA	ATCVLVPLG	VPVAWVLARL	AFPGRALVLR	LLMLPFVMP
	101	LVAGVGVLLAL	FGADGLLWRG	RQDTPYLLLY	GNVFFNLPVL	VRAAYQGFAQ
	151	VPAARLQTAR	TLGAGAWRPF	WDIEMPVLRP	WLAGGVCLVF	LYCFSGFGLA
15	201	LLLGGSRAT	VEVEIYQLVM	FELDMAGASA	LVWLVLGVTA	AAGLLYAWFG
	251	RRAVSDKAVS	PVMPSPQSV	GEYVLLAFSV	AVLSVCCLEP	LSAIVVKAWS
	301	AGESRRVLME	SETWQAVWNT	LRFSAAAVFA	AAVLGVVYAA	AARRLVWMRG
	351	LVFLPEMVSP	VCVSAGVLLL	YPGWASLPL	LLAMYALLAY	PFVAKDVLSA
	401	WDALPPDYGR	AAAGLGANGF	QTACRITFPL	LKPALRRGLT	LAAATCVGEF
	451	AATFLSRPE	WQTLTTLIYA	YLGRAGEDNY	ARAMVLTLLL	SAFAVCIFLL
20	501	LDNDEGGKRT	ETL*			

Further work revealed a variant gonococcal DNA sequence <SEQ ID 581>:

	1	ATGGATGGAC	GGTGTGGGC	GGTACGGGT	GCTTTTCCC	TGCTGCCTTC
	51	GGCTTTTTG	GCGGTAATGG	TCGTGCGCC	TTGTGGGCG	GTGGCGCGT
25	101	ATGACGGTTT	GGCGTGGCG	GCGGTGCTGT	CGGATGCCTA	TATGCTCAAA
	151	CGTTTGGCGT	GGACGGTGT	TCAGGCGCG	GCAACCTGTG	TGCTGGTGT
	201	GCCTTGGGC	GTGCTGTCG	CGTGGGTGCT	GGCGCGGCTG	GCGTCCCGG
	251	GGCGGCTTT	GGTGTGCGC	CTGTGATGC	TGCCGTTTGT	GATGCCACG
	301	CTGGTGGCGG	GCGTGGCGT	GCTGCTCTG	TTCGGGGCGG	ACGGGTGTT
	351	GTGGCGCGG	CGGCAGGATA	CGCCGTATCT	GTTGTGTAC	GGCAATGTG
30	401	TTTTCAACCT	GCCCGTGTG	GTCAGGCGG	CGTATCAGG	GTTTGTCAA
	451	GTGCTGCGG	CACGGCTTCA	GACGGCACGG	ACGTTGGGCG	CGGGGCGTG
	501	GCGGCGGTTT	TGGGACATG	AAATGCCCGT	TTTGCGCCCG	TGGCTGCCG
	551	GCGGCGGTG	CCTGTCTTC	CTGTATTGT	TTTGGGGTT	CGGGTGGCA
	601	TTGCTGTTG	GCGGCGCGG	TTATGCCACG	GTCGAAGTGG	AAATTTACCA
35	651	GTTGGTTATG	TTCGAACGCG	ATATGGCGG	GGCTTCGGCG	CTGGTGGCG
	701	TGGTGTGGG	GGTAACGCG	GCGGCGGGT	TGCTGTATGC	GTGGTTCGGC
	751	AGGCGCGCG	TTTCGGATAA	GGCGGTTCC	CCCGTGATGC	CGTCGCGCC
	801	GCAATCGGTG	GGGAATATG	TATTGCTGG	ATTTTCGGTG	GCGGTGTGT
	851	CCGTGTGCTG	CCTGTTTCT	TTGTGCGCA	TTGTTGTGAA	AGCGTGGTCG
40	901	GCCGCGAAT	CGCGCGGTG	GTTAATGGAA	AGTGAACGT	GGCAGGCAGT
	951	GTGGAATACT	ttGCGCTTT	CGGCGCGCG	GGTGTTCG	GCGGCGGTTT
	1001	TGGGTGTGGT	GTATGCGCG	GCGGCGCGG	GGCTGGTGTG	GATGCGCGGA
	1051	CTGGTGTGTT	TACGTTTAT	GGTGTGCGG	GTTTGTGTT	CGGCGGCGT
	1101	GCTGCTGCTT	TATCCGGGT	GGACGGCTTC	GTTACCGCTG	CTGCTGGCGA
45	1151	TGTATGCGCT	GCTGGCGTAT	CCGTGTGTG	CAAAAGATGT	TTTATCGGCC
	1201	TGGGATGCAC	TGCCGCGGA	TTACGGCAGG	GCGGCGGCAG	GTTTGGGCGC
	1251	AAACGGCTTT	CAGACGGCAT	GCCGTATCAC	GTTCCCTCTC	TTGAAACCGG
	1301	CGTTGCGCG	CGGTCTGACT	TTGGCGGCG	CGACGTGTGT	GGGCGAATTT
	1351	GCGGCAACCT	TGTTCTGTCT	GCGTCCGGA	TGGCAGACGT	TGACGACTTT
50	1401	GATTTATGCC	TATTTGGGG	GTGCGGGTGA	GGACAATTAT	GCGGCGGCAA
	1451	TGGTGTGAC	ATTGCTGTG	TCGGCATTG	CGGTGTGCAT	TTTCTGCTG
	1501	TTGGACAACG	GCGAAGGCG	aaaACGGACG	GAAACGTTAT	AA

This corresponds to the amino acid sequence <SEQ ID 582; ORF139ng-1>:

55	1	MDGRCWAVRG	AFSLLPSAFL	AVMVVAPLWA	VAAYDGLAWR	AVLSDAYMLK
	51	RLAWTVFQAA	ATCVLVPLG	VPVAWVLARL	AFPGRALVLR	LLMLPFVMP
	101	LVAGVGVLLAL	FGADGLLWRG	RQDTPYLLLY	GNVFFNLPVL	VRAAYQGFAQ
	151	VPAARLQTAR	TLGAGAWRPF	WDIEMPVLRP	WLAGGVCLVF	LYCFSGFGLA
	201	LLLGGSRAT	VEVEIYQLVM	FELDMAGASA	LVWLVLGVTA	AAGLLYAWFG
	251	RRAVSDKAVS	PVMPSPQSV	GEYVLLAFSV	AVLSVCCLEP	LSAIVVKAWS
	301	AGESRRVLME	SETWQAVWNT	LRFSAAAVFA	AAVLGVVYAA	AARRLVWMRG
	351	LVFLPEMVSP	VCVSAGVLLL	YPGWASLPL	LLAMYALLAY	PFVAKDVLSA
	401	WDALPPDYGR	AAAGLGANGF	QTACRITFPL	LKPALRRGLT	LAAATCVGEF
	451	AATFLSRPE	WQTLTTLIYA	YLGRAGEDNY	ARAMVLTLLL	SAFAVCIFLL
60	501	LDNDEGGKRT	ETL*			

ORF139ng-1 and ORF139-1 show 95.9% identity over 513aa overlap:

```

5      orf139ng      MDGRCWAVRGAFSLLPSAFLAVMVVAPLWAVAAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
      orf139-1      MDGRRWVWVGAFALLPSAFLAVMVVAPLWAVAAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
      orf139ng      ATCVLVLPPLGVFPAWVLARLAFPGRALVLRLLMLPFVMPTLVAGVGVLALFGADGLLWRG
      orf139-1      ATCVLVLPPLGVFPAWVLARLAFPGRALVLRLLMLPFVMPTLVAGVGVLALFGADGLLWRG
10     orf139ng      RQDTPYLLLYGNVFFNLPLVLVRAAYQGFAQVPAARLQTARTLGAGAWRRFWDIEMPVLRP
      orf139-1      RQDTPYLLLYGNVFFNLPLVLVRAAYQGFVQVPAARLQTARTLGAGAWRRFWDIEMPVLRP
      orf139ng      WLAGGVCLVFLYCFSGFGLALLLGGSRATVEVEIYQLVMFELDMAGASALVWLVLGVTA
15     orf139-1      WLAGGVCLVFLYCFSGFGLALLLGGSRATVEVEIYQLVMFELDMAGASALVWLVLGVTA
      orf139ng      AAGLLYAWFGRRRAVSDKAVSPVMPSPQSVGEYVLLAFSAVLSVCCLFPLSAIVVKAWS
20     orf139-1      AAGLLYAWFGRRRAVSDKAVSPVMPSPQSVGEYVLLAFSAVLSVCCLFPLSAIVVKAWS
      orf139ng      AGESRRVLMESETWQAVWNTLRFSAAAVFAAAVLGVVYAAAARRLVWMRGLVFLPFMVSP
      orf139        AGESWRVLMESETWQAVWNTLRFSAAAVYAAAVLGVVYAAAARRSAWMRGLMFLPFMVSP
25     orf139ng      VCVSAGVLLLYPGWTASLPLLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAGLGANGF
      orf139-1      VCVSAGVLLLYPGWTASLPLLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAGLGANGF
30     orf139ng      QTACRITFPLLKPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY
      orf139-1      QTACRITFPLLKPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY
35     orf139ng      ARAMVLTLLLSAFAVCIFLLLDNGEGGKRTETL
      orf139-1      ARAMVLTLLLAFAFGIFLLLDGGEGGKQTETL

```

Based on the presence of a predicted binding-protein-dependent transport systems inner membrane component signature (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 70

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 583>:

```

45      1  ATGGACGGCT  GGACACAGAC  GCTGTCCGCG  CAAACCCTGT  TGGGCATTTC
      51  GCGCGCGGCA  ATCATCTCA  TTCTGATTTT  AATCGTCAGA  TTCCGCATCC
      101  ACGCGCTGCT  GACACTGGTC  ATCGTCAGCC  TGCTGACGGC  TTTGGCAACC
      151  GGTTCGCCCA  CAGGCAGCAT  TGTCAAAGAC  ATACTGGTCA  AAAACTTCGG
      201  CGGCACGCTC  GCGCGCTGG  CGCTTCTGGT  CGGCCTGGGC  GCGATGCTCG
      251  AACGTTTGGT  C...

```

This corresponds to the amino acid sequence <SEQ ID 584; ORF140>:

```

50      1  MDGWTQTLA  QTLGISA  IILILILIVR  FRIHALTLV  IVSLLTALAT
      51  GLPTGSIVKD  ILVKNFG  GTL  GGVALLVGLG  AMLERLV..

```

Further work revealed the complete nucleotide sequence <SEQ ID 585>:

```

55      1  ATGGACGGCT  GGACACAGAC  GCTGTCCGCG  CAAACCCTGT  TGGGCATTTC
      51  GCGCGCGGCA  ATCATCTCA  TTCTGATTTT  AATCGTCAAA  TTCCGCATCC
      101  ACGCGCTGCT  GACACTGGTC  ATCGTCAGCC  TGCTGACGGC  TTTGGCAACC
      151  GGTTCGCCCA  CAGGCAGCAT  TGTCAAAGAC  ATACTGGTCA  AAAACTTCGG

```

-334-

201 CGGCACGCTC GCGGCGGTGG CGCTTCTGGT CGGCCTGGGC GCGATGCTCG  
 251 GACGTTTGGT CGAAACATCC GCGGCGCAC AGTCGCTGGC GGACGCGCTG  
 301 ATCCGGATGT TCGGCGAAAA ACGCGCACCG TTCGCGCTGG GCGTTGCCTC  
 351 GCTGATTTTC GGCTTCCCGA TTTTCTTCGA TGCCGGACTA ATCGTCATGC  
 401 TGCCCATCGT GTTCGCCACC GCACGGCGCA TGAAACAGGA CGTACTGCCC  
 451 TTCCGCGTTG CCTCCATCGG CGCATTTTCC GTCATGCACG TCTTCCTGCC  
 501 GCCCATCCG GCGCCGATTG CCGCTTCCGA ATTTTACGGC GCGAACATCG  
 551 GCCAAGTTTT GATTTTGGGT CTGCCGACCG CCTTCATCAC ATGGTATTTT  
 601 AGCGGCTATA TGCTCGCAA AGTGTTGGGG CGCACCATCC ATGTTCCTGT  
 651 TCCCGAATG CTCAGCGCG GCACGCAAGA CAACGACCTG CCGAAAGAAC  
 701 CTGCCAAAGC AGGAACGGTC GTCGCCATCA TGCTGATCC CATGCTGCTG  
 751 ATTTTCTGA ATACCGCGGT ATCGGCCCTC ATCAGCGAAA AACTCGTAAG  
 801 TCGGACGAA ACCTGGGTT AGACGGCAA AATAATCGGT TCGACACCGA  
 851 TCGCCCTTCT GATTTCCGTA TTGGTCGCAC TGTTTGTCTT GGGACGCAA  
 901 GCGGCGAAA GCGCGACGCG GTTGGAAAA ACCGTGGACG GCGCACTCGC  
 951 CCCGCTCTGT TCCGTGATTG TGATTACCGG CGCGGGCGGT ATGTTCGGCG  
 1001 GCGTTTTCG CGCTTCCGGC ATCGGCAAGG CACTCGCCGA CAGCATGGCG  
 1051 GATTTGGGCA TTCCCGTCTT TTTGGGCTGT TTCCTTGTCT CCTTGGCACT  
 1101 GCGTATCGCG CAAGGTTTCG CAACCGTCGC CCTGACCACC GCCGCGCGCG  
 1151 TGATGGCTCC TGCCGTGGC GCCGCGGCT TTACCGACTG GCAGCTCGCC  
 1201 TGTATCGTAT TGGCAACGGC GGCAGGTTCT GTCGGTTGCA GCCACTTCAA  
 1251 CGACTCCGGC TTCTGGCTGG TCGCCGTCT CTTGGACATG GACGTACCGA  
 1301 CCACGCTGAA AACCTGGACG GTCAACAAA CCCTCATCGC ACTCATCGGC  
 1351 TTTGCCTTGT CCGCACTGCT GTTCGCCATC GTCTGA

25 This corresponds to the amino acid sequence <SEQ ID 586; ORF140-1>:

1 MDGWTQTLA QTLGISAAA IILILILIVK FRIHALTLV IVSLLTALAT  
 51 GLPTGSIVND ILVKNFGGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL  
 101 IRMFGEKRAP FALGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVLV  
 151 FALASIGAFS VMHVLPHP GPIAASEFYG ANIGQVLILG LPTAFITWYF  
 201 SGYMLGKVLG RTIHVPVPEL LSGGTQDNDL PKEPAKAGTV VAIMLIPMLL  
 251 IFLNTGVSAL ISEKLVSAD TWVQTAKIIG STPIALLISV LVALEVLGRK  
 301 RGESESALEK TVDGALAPVC SVLITGAGG MFGGVLRASG IGKALADMSA  
 351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA  
 401 CIVLATAAGS VGSFHNDSG FWLVGRLLDM DVPTTLKTWT VNQTLIALIG  
 451 FALSALLFAI V\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF140 shows 95.4% identity over a 87aa overlap with an ORF (ORF140a) from strain A of *N. meningitidis*:

40	orfl40.pep	MDGWTQTLAQTLLGISAAAIIILILILIVREFRIHALTLVIVSLLTALATGLPTGSIVKD
	orfl40a	MDGWTQTLAQTLLGISAAAIIILILILIVREFRIHALTLVIVSLLTALATGLPTGSIVND
45	orfl40.pep	ILVKNFGGTLGGVALLVGLGAMLERLV
	orfl40a	VLVKNFGGTLGGVALLVGLGAMLRVETSGGAQSLADALIRMFGEKRAPFALGVASLIF

The complete length ORF140a nucleotide sequence <SEQ ID 587> is:

1 ATGGACGGCT GGACACAGAC GCTGTCCGCG CAAACCCTGT TGGGCATTTT  
 51 GCGGCGGCA ATCATCTCA TTCTGATTTT AATCGTCAA TTCCGCATCC  
 101 ACGCGCTGCT GACACTGGTC ATCGTCAGCC TGCTGACGGC TTTGGCAACC  
 151 GGTTTGCCCA CAGGCAGCAT TGTCACGAC GACTGCTGCA AAAACTTCGG  
 201 CGGCACGCTC GCGGCGGTGG CGCTTCTGGT CGGCCTGGGC GCGATGCTCG  
 251 GACGTTTGGT CGAAACATCC GCGGCGCAC AGTCGCTGGC GGACGCGCTG  
 301 ATCCGGATGT TCGGCGAAAA ACGCGCACCG TTCGCGCTGG GCGTTGCCTC  
 351 GCTGATTTTC GGCTTCCCGA TTTTCTTCGA TGCCGGACTA ATCGTCATGC  
 401 TGCCCATCGT GTTCGCCACC GCACGGCGCA TGAAACAGGA CGTACTGCCC  
 451 TTCGCGCTTG CCTCCATCGG CGCATTTTCC GTCATGCACG TCTTCCTGCC

-335-

501 GCCCATCCG GGCCGATTG CCGCTTCCGA ATTTTACGGC GCGAACATCG  
 551 GCCAAGTTT GATTTTGGGT CTGCCGACCG CCTTCATCAC ATGGTATTTT  
 601 AGCGGCTATA TGCTCGGCAA AGTGTGGGG CGCACCATCC ATGTTCCCGT  
 651 TCCCGAACTG CTCAGCGGCG GCACGCAAGA CAACGACCTG CCGAAAGAAC  
 5 701 CTGCCAAAGC AGGAACGGTC GTCGCCATCA TGCTGATTCC CATGCTGCTG  
 751 ATTTTCTGA ATACCGGCGT ATCGGCCCTC ATCAGCGAAA AACTCGTAAG  
 801 TGCGGACGAA ACCTGGGTTC AGACGGCAA AATAATCGGT TCGACACCGA  
 851 TCGCCCTTCT GATTTCGTA TTGGTCGCAC TGTGTGTCTT GGGACGCAAA  
 901 CGCGGCGAAA GCGGCAGCGC GTTGAAAAA ACCGTGGACG GCGCACTCGC  
 10 951 CCCGCTCTGT TCCGTGATTG TGATTACCGG CGCGGGCGGT ATGTTCCGGC  
 1001 GCGTTTTGCG CGCTTCCGGC ATCGGCAAGG CACTCGCCGA CAGCATGGCG  
 1051 GATTTGGGCA TTCCCGTCCT TTTGGGCTGT TTCCTGTGCG CCTTGGCACT  
 1101 GCGTATCGCG CAAGGTTCCG CAACCGTCGC CCTGACCACC GCCGCCGCGC  
 1151 TGATGGCTCC TGCCGTGGC GCCGCCGGCT TTACCGACTG GCAGCTCGCC  
 15 1201 TGTATCGTAT TGGCAACGGC GGCAGGTCG GTCGGTTGCA GCCACTTCAA  
 1251 CGACTCCGGC TTCTGGCTGG TCGGCCGCCT CTTGGACATG GACGTACCGA  
 1301 CCACGCTGAA AACCTGGACG GTCAACCAA CCCTCATCGC ACTCATCGGC  
 1351 TTTGCCTTGT CCGCACTGCT GTTCGCCATC GTCTGA

This encodes a protein having amino acid sequence <SEQ ID 588>:

20 1 MDGWTQTLQA QTLGISAAA IILILILIVK FRIHALLTLV IVSLLTALAT  
 51 GLPTGSIVND VLVKNFGGTL GGVALLVGLG AMLGRIVETS GGAQSLADAL  
 101 IRMFGEKRAP FALGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVL  
 151 FALASIGAFS VMHVFLPPHP GPIAASEFYG ANIGQVLILG LPTAFITWYF  
 201 SGYMLGKVLG RTIHVPVPEL LSGGTQDNDL PKEPAKAGTV VAIMLIPMLL  
 25 251 IFLNTGVSAL ISEKLVSAD E TWVQTAKIIG STPIALLISV LVALFVLGRK  
 301 RGESESALEK TVDGLAPVC SVILITGAGG MFGGVLRSAG IGKALADSM  
 351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA  
 401 CIVLATAAGS VGCSEFNDSE FWLVGRLLDM DVPTTLKTWT VNQTLIALIG  
 451 FALSALLFAI V\*

30 ORF140a and ORF140-1 show 99.8% identity over a 461aa overlap:

orfl40-1.pep MDGWTQTLQAQTLGISAAAIILILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND 60  
 orfl40a MDGWTQTLQAQTLGISAAAIILILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND 60  
 35 orfl40-1.pep ILVKNFGGTLGGVALLVGLGAMLGRIVETS GGAQSLADALIRMFGEKRAPFALGVASLIF 120  
 orfl40a ILVKNFGGTLGGVALLVGLGAMLGRIVETS GGAQSLADALIRMFGEKRAPFALGVASLIF 120  
 40 orfl40-1.pep GFPIFFDAGLIVMLPIVFATARRMKQDVLFPFALASIGAFSVMHVFLPPHPGPIAASEFYG 180  
 orfl40a GFPIFFDAGLIVMLPIVFATARRMKQDVLFPFALASIGAFSVMHVFLPPHPGPIAASEFYG 810  
 orfl40-1.pep ANIGQVLILGLPTAFITWYFSGYMLGKVLGRRTIHVPVPELLSGGTQDNDLPKEPAKAGTV 240  
 45 orfl40a ANIGQVLILGLPTAFITWYFSGYMLGKVLGRRTIHVPVPELLSGGTQDNDLPKEPAKAGTV 240  
 orfl40-1.pep VAIMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKIIGSTPIALLISV LVALFVLGRK 300  
 orfl40a VAIMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKIIGSTPIALLISV LVALFVLGRK 300  
 50 orfl40-1.pep RGESESALEKTV DGLAPVCSVILITGAGGMFGGVLRSAGIGKALADSMADLGIPVLLGC 360  
 orfl40a RGESESALEKTV DGLAPVCSVILITGAGGMFGGVLRSAGIGKALADSMADLGIPVLLGC 360  
 55 orfl40-1.pep FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSEFNDSE 420  
 orfl40a FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSEFNDSE 420  
 60 orfl40-1.pep FWLVGRLLDM DVPTTLKTWTVNQTLIALIGFALSALLFAIV 461  
 orfl40a FWLVGRLLDM DVPTTLKTWTVNQTLIALIGFALSALLFAIV 461

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF140 shows 92% identity over a 87aa overlap with a predicted ORF (ORF140ng) from

65 *N.gonorrhoeae*:

10	1	<u>MDGRTOTLSA</u>	<u>QTLGISAAA</u>	<u>IILILILIVK</u>	<u>FRIRALLTLV</u>	<u>IASLLTALAT</u>
	51	<u>GLPTGSIVND</u>	<u>VLVKNFEGTL</u>	<u>GGVALLVGLG</u>	<u>AMLGRLVETS</u>	<u>GGAQSLADAL</u>
	101	<u>IRMFGEKRAP</u>	<u>FAPGVASLIF</u>	<u>GFPIFFDAGL</u>	<u>IVMLPIVFAT</u>	<u>ARRMKQDVLP</u>
	151	<u>FALASVGAFS</u>	<u>VMHVFLPPHP</u>	<u>GPAASEFYG</u>	<u>ANIGQVLILG</u>	<u>LPTAFITWYF</u>
15	201	<u>SGYMLGKVLG</u>	<u>RAIHVPPEL</u>	<u>LSGGTQSDMP</u>	<u>PKPEKAGTV</u>	<u>VAVMLIPMLL</u>
	251	<u>IFLNTGVSAL</u>	<u>ISEKLVSAD</u>	<u>TWVQTAKMIG</u>	<u>STPEPALLISV</u>	<u>LAALLVLGRK</u>
	301	<u>RGESGSTLEK</u>	<u>TVDGALAPAC</u>	<u>SVILITGAGG</u>	<u>MFGGVLRASG</u>	<u>IGKALADSM</u>
	351	<u>DLGIPVLLGC</u>	<u>FLVALALRIA</u>	<u>QGSATVALTT</u>	<u>AAALMAPAVA</u>	<u>AAGFTDWQLA</u>
	401	<u>CIVLATAAGS</u>	<u>VGCSHFNDSG</u>	<u>FWLVGRLSDM</u>	<u>DVPTTLKTTW</u>	<u>VNQTLLIAFIG</u>
	451	<u>FALSALLFAI</u>	<u>V*</u>			

	1	ATGGACGGCC	GGACACAGAC	GCTGTCCGCG	CAAACCTTGT	TGGGCATTTT
	51	GGCGGCGGCA	ATCATCTCTCA	TTCTGATTTT	AATCGTCAAA	TTCCGCATCC
	101	GCGCGCTGCT	GACACTGGTC	ATCGCCAGCC	TGCTGACGGC	TTTGGCAACC
25	151	GGTTTGCCCA	CAGGCAGCAT	CGTCAACGAC	GTACTGGTCA	AAAACCTTCGG
	201	CGGCACGCTC	GGCGGCGTGG	CGCTTCTGGT	CGGCTCTGGG	GCAATGCTCG
	251	GACGTTTGGT	AGAAACATCC	GGCGGCGCAC	AGTCGCTGGC	GGACGCGCTG
	301	ATCCGGATGT	TCGGCGAAAA	ACGCGCACCG	TTGCTCCGG	CGGTTGCCTC
	351	GCTGATTTTC	GGCTTCCCGA	TTTTCTTCGA	TGCCGGACTA	ATCGTCATGC
30	401	TGCCCATCGT	ATTGCGCACC	GCACGGCGCA	TGAACAGGA	CGTACTGCCC
	451	TTCCGCGTTG	CCTCCGTCGG	CGCATTTTCC	GTCATGCACG	TCTTCTGCCC
	501	GCCCCATCCG	GGCCCCGATTG	CCGCTTCCGA	ATTTTACGGC	GCGAACATCG
	551	GCCAGTTTTT	GATTTTGGGT	CTGCCGACCG	CCTTCATCAC	ATGGTATTTT
	601	AGCGGCTATA	TGCTCGGCAA	AGTGTGGGG	CAGCGCATTC	ATGTTCCCGT
35	651	TCCCGAATCG	CTCAGCGGCG	GCACGCAAGA	CAGCGACCCG	CCGAAAGAAC
	701	CTGCCAAAGC	AGGAACGGTC	GTCGCCGTCA	TGCTGATTCC	CATGCTGCTG
	751	ATTTTCTCTGA	ATACCGCGGT	ATCAGCCCTC	ATCAGCGAAA	AACTCGTAAG
	801	TGCGGACGAA	ACTTGGGTTT	AGACGGCAAA	AATGATCGGT	TCGACACCTG
	851	TGCCCTTCT	GATTTCCGTA	TTGGCCGCAC	TGTTGGTCTT	GGGACGCAAA
40	901	CGCGCGGAAA	GCGGCAGCAC	GTTGGAAAAA	ACCGTGGACG	CGCGCATTCG
	951	CCCCGCTGT	TCCGTGATT	TGATTACCGG	CGCGGGCGGT	ATGTTCCGGC
	1001	GCGTTTTCGG	CGCTTCCGGC	ATCGGCAAGG	CACTCGCCGA	CAGCATGGCG
	1051	GATTTGGGCA	TTCCCGTCTT	TTTGGGCTGC	TTCTTGTGTC	CCTTGGCACT
	1101	CGGTATCGCG	CAAGGTTTCG	CAACCGTCGC	CTGACCACA	CGCCCGCGCG
45	1151	TGATGGCTCC	TGCCGTTGCC	GCCGCGCGCT	TTACCGACTG	GCAGCTCGCC
	1201	TGTATCGTAT	TGGCAACGGC	GGCAGGTTTC	GTCGGTTGCA	GCCACTTCAA
	1251	CGACTCCGGC	TTCTGGCTGG	TCGGCCCGCT	CTTGGATATG	GACGTACCGA
	1301	CCACGCTGAA	AACCTGGACG	GTCACACCAA	CCCTCATCGC	ATTCATCGGC
	1351	TTTGCTTGT	CCGCACTGCT	GTTTGGCATC	GTCGTA	

50	1	MDGRTOTLSA	QTLLGISA	AAILILILIVK	FRIRALLTLV	IASLLTALAT
	51	GLPTGSIVND	VLVNKFGGTL	GGVALLVGLG	AMLGRLVETS	GGAQSLADAL
	101	IRMFEKRAP	FAPGVASLIF	GPIFFDAGL	IVMLPIVFAT	ARRMKQDVLP
	151	FALASVGAFS	VMHVFLPPHP	GPIAASEFYG	ANIGQVLIIG	LPTAFITWYF
	201	SGYMLGKVLG	RAIHVPPEL	LSGGTQSDP	PKEPAKAGTV	VAVMLIPLML
55	251	IFLNTGVSAL	ISEKLVSAD	TWVQTAKMIG	STPVALLISV	LAALLVLGRK
	301	RGESGSTLEK	TVDGALAPAC	SVILITGAGG	MFGGVLRASG	IGKALADSMA
	351	DLGIPVLLGC	FLVALALRIA	QGSATVALTT	AAALMAPAVA	AAGFTDWQLA
	401	CIVLATAAGS	VGCSEHND	SGFWLVGRLLDM	DVPTTLKTTWT	VNQTLLIAFIG
	451	FALSALLFA	V*			

orf140ng-1.pep MDGRTQTLSAQTLGLISAAAIIILILIVKFRIRALLTLVIASLLTALATGLPTGSIVND  
||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

-337-

```

orfl40-1      MDGWTQTLAQTLGISAAILILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND
orfl40ng-1.pep VLVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFG EKRAPFAPGVASLIF
5 orfl40-1      ILVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFG EKRAPFALGVASLIF
orfl40ng-1.pep GFPIFFDAGLIVMLPIVFATARRMKQDVLPFALASVGAFSVMHVFLPPHPGPIAASEFYG
10 orfl40-1      GFPIFFDAGLIVMLPIVFATARRMKQDVLPFALASIGAFSVMHVFLPPHPGPIAASEFYG
orfl40ng-1.pep ANIGQVLILGLPTAFITWYFSGYMLGKVLGRAIHVPVPELLSGGTQSDSDPPKEPAKAGTV
orfl40-1      ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTIHVPVPELLSGGTQDNDLPKEPAKAGTV
15 orfl40ng-1.pep VAVMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKMIGSTPVALLISVLAALLVLGRK
orfl40-1      VAIMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKIIGSTPIALLISVLVALFVLGRK
orfl40ng-1.pep RGE SGSTLEKTVDGALAPACSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC
20 orfl40-1      RGE SGSALEKTVDGALAPVCSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC
orfl40ng-1.pep FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSHFND SG
25 orfl40-1      FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSHFND SG
orfl40ng-1.pep FWLVGRLLDMDVPTTLKTWTVNQTLIAFIGFALSALLFAIV
orfl40-1      FWLVGRLLDMDVPTTLKTWTVNQTLIALIGFALSALLFAIV

```

30 Furthermore, ORF140ng-1 is homologous to an *E.coli* protein:

```

gi|882633 (U29579) ORF_o454 [Escherichia coli] >gi|1789097 (AE000358) o454;
This 454 aa ORF is 34% identical (9 gaps) to 444 residues of an approx. 456 aa
protein GNTP BACLI SW: P46832 [Escherichia coli] Length = 454
Score = 210 bits (529), Expect = 1e-53
35 Identities = 130/384 (33%), Positives = 194/384 (49%), Gaps = 19/384 (4%)

Query: 88 ETSGGAQSLADALIRMFG EKRAPFAPGVASLIFGFPIFFDAGLIVMLPIVFATARRMKQD 147
E S GGA+SLA+ R G+KR A +A+ G P+FFD G I++ PI++ A+ K
40 Sbjct: 80 EHS GGAESLAN YFSRKL GDKRTIAALT LA AFFLGIPVFFDVGFILAPIIYGF AKVAKIS 139

Query: 148 VLPFALASVGAFSVMHVFLPPHPGPIAASEFYGANIGQVLILGLPTAFITWYFSGYMLGK 207
L F L G +HV +PPHPG+AA+ A+IG + I+G+ + I GY K
Sbjct: 140 PLKFGLPVAGIMLT VHVAVPPHPGPVAAAGLLHADIGWLTIIIGIAIS-IPVG VGYFAAK 198

45 Query: 208 VLGRAIHVPVPELL-----SGGTQSDSDPPKEPAKAGTVVAVMLIPMLLIFLNTGV 257
++ + + E+L G T+ SD P A V ++++IP+ +I T
Sbjct: 199 IINKRQYAMSVEVLEQMQLAPASEEGATKLS DKINPPGVA-LVTS LIVIPIAIIMAGT-- 255

50 Query: 258 SALISEKLVSADETWVQTAKMIGSTPXXXXXXXXXXXXXGRKRGE SGSTLEKTVDGALA 317
+S L+ + T ++IGS +RG S + AL
Sbjct: 256 ---VSATLMPPSHPLGLTLQLIGSPMVALMIALVLAFWLLALRRGWSLQHTSDIMGSALP 312

Query: 318 PACSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGCFLVALALRIAQGSXXXXX 377
A VIL+TGAGG+FG VL SG+GKALA+ + + +P+L F+++LALR +QGS
55 Sbjct: 313 TAAVVILVTGAGGVFGKVLVESGVGKALANMLQ MIDLP LLPAAFIISLALRASQGS--AT 370

Query: 378 XXXXXXXXXXXXXXXXGFTDWQLACIVLATAAGSVGCSHFND SGFWLVGRLLDMDVPTTLK 437
G Q + LA G +G SH NDSGFW+V + L + V LK
60 Sbjct: 371 VAILTTGGLLSEAVMGLNPIQCVLVTLAACFGGLGASHINDSGFWIVTKYLGLSVADGLK 430

Query: 438 TWTVNQTLIAFIGFALSALLFAIV 461
TWTV T++ F GF ++ ++A++
Sbjct: 431 TWTVLTTLIGFTGFLITWCVWAVI 454

```

Based on this analysis, including the identification of the presence of a putative leader sequence

65 (double-underlined) and several putative transmembrane domains (single-underlined) in the

gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 71

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 593>:

```

5      1  ..GATTTCGGCA TATCGCCCGT GTATCTTTGG GTTGCCGCGC CGTTCAAACA
      51  TTTGCTGTCT CCGTGGGCTG CCGACTCATA CGATGTCGCA CGCTTTGCAG
     101  GCGTATTTTT TGCCGTTATC GGACTGACTT CCTGCGGCTT TGCCGGTTTC
     151  AACTTTTTTG GCAGACACCA CGGGCGCAC. GTCGTCCTGA TTCTCATCGG
     201  CTGTATCGGG CTGATTCCAG TTGCCCATTT CCTCAACCCC GCTGCCGCCG
     251  CCTTTGCCCG CGCCGGACTG GTGCTGCACG GTTATTCTTT GGCTCGCCGG
     301  CGCGTGATTG CCGCTCTTT TCTGCTCGGT ACGGGCTGGA CGCTGATGTC
     351  GTTGGCAGCA GCTTATCCGG CAGCATTTGC CCTGATGCTG CCCTTGCCCC
     401  TACTGATGTT TTTCCGTCCG ..

```

This corresponds to the amino acid sequence <SEQ ID 594; ORF141>:

```

15      1  ..DFGISPVYLW VAAAFKLLS PWAADSYDVA RFAGVFFAVI GLTSCGFAGF
      51  NFLGRHHGRX VVLILIGCIG LIPVAHFLNP AAAFAAAGL VLHGYSLARR
     101  RVIAASFLLG TGWTLMSLAA AYPAAFALML PLPVLMEFRP ..

```

Further work revealed the complete nucleotide sequence <SEQ ID 595>:

```

20      1  ATGCTGACCT ATACCCCGCC CGATGCCCGC CCGCCCGCCA AAACCCACGA
      51  AAAGCCGTGG CTGCTGCTGT TGATGGCGTT TGCCTGGTTG TGGCCCGCGC
     101  TGTTTTCCCA CGATTGTGGG AATCTGACG AACCTGCCGT CTATACCGCC
     151  GTCGAAGCAC TGGCAGGCAG CCCACCCCC TTGTTGCCCC ATCTGTTCCG
     201  TCAAACCGAT TTCGGCATA CCGCCGTGTA TCTTTGGGTT GCCGCCCGCT
     251  TCAAACATTT GCTGTGCGCG TGGGTGCGG ACTCATACGA TGCCGCACGC
     301  TTTGCAGGCG TATTTTTTGC CGTTATCGGA CTGACTTCCT GCGGCTTTGC
     351  CGGTTTCAAC TTTTGGGCA GACACCACGG GCGCAGCGTC GTCCTGATTC
     401  TCATCGGCTG TATCGGGCTG ATTCCAGTTG CCCATTTCCT CAACCCCGCT
     451  GCGCCGCCTT TTGCCCGCGC CGGACTGGTG CTGCACGGTT ATTCTTTGGC
     501  TCGCCGCGCG GTGATTGCCG CCTCTTTTCT GCTCGGTACG GGCTGGACGC
     551  TGATGTCGTT GGCAGCAGCT TATCCGGCAG CATTTGCCCT GATGCTGCCC
     601  TTGCCCGTAC TGATGTTTTT CCGTCCGTGG CAAAGCAGGC GTTTGATGTT
     651  GACGGCAGTC GCCTCACTTG CCTTTGCCCT GCCGCTTATG ACCGTTTACC
     701  CGCTGCTCTT GGCAAAAACG CAGCCCGCGC TGTTGCGCGA ATGGCTCGAC
     751  TATCAGTTT TCGGTACGTT CGGCGGCGTG CGGCACGTT AGACGGCATT
     801  CAGTTTGTTT TACTATCTGA AAAACCTGCT TTGTTTGCA TTGCCCGCGC
     851  TGCCGCTGGC GGTTTGGACG GTTTGCCGCA CGCGCTGTT TTCGACCGAC
     901  TGGGGGATTT TGGGCGTCGT CTGGATGCTT GCCGTTTTGG TGCTGCTTGC
     951  CGTCAATCCG CAGCGTTTTT AGGATAACCT CGTCTGGCTG CTTCCGCCGC
    1001  TTGCCCTGTT CGGCGCGGCG CAACTGGACA GCCTGAGGCG CGGCGCGGCG
    1051  CGGTTTGTC ACTGTTTCGG CATTATGGCG TTCGGACTGT TTGCCGTGTT
    1101  CCTGTGGACG GGCTTTTTTC CCATGAATTA CGGCTGGCCC GCCAAGCTTG
    1151  CCGAACGCGC CGCTATTTC AGCCCGTATT ATGTTCTCTA TATCGATCCC
    1201  ATTCGATGCG CGGTGCGCT ACTGTTTACA CCCTTGTTGG TGTGGGCGAT
    1251  TACCCGAAAA AACATACGCG GCAGGCAGGC GGTACCAAC TGGGCGGCAG
    1301  GCGTTACCCT GACCTGGGCT TTGCTGATGA CGCTGTTTCT GCCGTGGCTG
    1351  GACGCGCGCA AAAGCCACGC GCCGTCGTC CGGAGTATGG AGGCATCGCT
    1401  TTCCCGGAA TTGAAACGGG AGCTTTCAGA CGGCATCGAG TGTATCGCA
    1451  TAGGCGCGCG CGACCTGCAC ACGCGATTG TTTGGACGCA GTACGGCACA
    1501  TTGCCGCACC GCGTCGGCGA TGTACAATGC CGCTACCGCA TCGTCTCCT
    1551  GCCCCAAAT GCGGATGCGC CGCAAGGCTG GCAGACGGTT TGGCAGGGTG
    1601  CGCGTCCGCG CAACAAAGAC AGTAAGTTCT CACTGATACG GAAATCGGG
    1651  GAAATATAT AA

```

This corresponds to the amino acid sequence <SEQ ID 596; ORF141-1>:

```

55      1  MLTYTPPDAR PPAKTHEKPW LLLMAFAWL WPGVFSHDLW NPDEPAVYTA
      51  VEALAGSPTP LVAHLFGQTD FGIPPVYLWV AAFKHLSP WAADSYDAAR
     101  FAGVFFAVIG LTSCGFAGFN FLGRHHGRSV VLILIGCIGL IPVAHFLNPA
     151  AAAFAAAGLV LHGYSLARRR VIAASFLLGT GWTLMSLAAA YPAAFALMLP
     201  LPVLMFFRPW QSRRLMLTAV ASLAFALPLM TVYPLLLAKT QPALFAQWLD

```

-339-

5

```

251 YHVFGEFGGV RHVQTAFLSF YYLKNLLWFA LPALPLAVVT VCRTLRFSTD
301 WGILGVVWML AVLVLAVNP QRFQDNLVWL LPPLALFGAA QLDLRRGAA
351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAFY SPYYVPDIDP
401 IPMAVAVLFT PLWLWAIKTRK NIRGRQAVTN WAAGVTLTWA LLMTLFLPWL
451 DAAKSHAPVV RSMEASLSPE LKRELSDGIE CIGIGGGDLH TRIVWTQYGT
501 LPHRVGDVQC RYRIVLLPQN ADAPOGWQTV WQGARPRNKD SKFALIRKIG
551 ENI*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

- 10 ORF141 shows 95.0% identity over a 140aa overlap with an ORF (ORF141a) from strain A of *N. meningitidis*:

```

                                10      20      30
orfl41.pep                    DFGISPVYLWVAAAFKHLLSPWAADSYDVA
15 orfl41a                    WNPDEPAVYTAVEALAGSPTPLVAHLFGQIDFGIPPVYLWVAAAFKHLLSPWAADPYDAA
                                40      50      60      70      80      90

                                40      50      60      70      80      90
20 orfl41.pep                RFAGVFFAVIGLTS CGFAGFNFLGRHHRXVVLILIGCIGLIPVAHFLNPAAAAFAAAGL
orfl41a                        RFAGVFFAVVGLTSCGFAGFNFLGRHHRSVVLILIGCIGLIPTVHFLNPAAAAFAAAGL
                                100     110     120     130     140     150

                                100     110     120     130     140
25 orfl41.pep                VLHGYSLARRRVIAASFLLGTGWTLSLAAAYPAAFALMLPLPVLMMFFRP
orfl41a                        VLHGYSLARRRVIAASFLLGTGWTLSLAAAYPAAFALMLPLPVLMMFFRPWQSRRLMLTA
                                160     170     180     190     200     210

30 orfl41a                    VASLAFALPLMTVYPLLLAKTOPALFAQWLDHVFGEFGGVRIQTAFSLFYLLKNLLWF
                                220     230     240     250     260     270

```

The complete length ORF141a nucleotide sequence <SEQ ID 597> is:

```

1  ATGCTGACCT ATACCCCGCC CGATGCCCGC CCGCCCGCCA AAACCCACGA
51 AAAGCCGTGG CTGTTGCTGT TGATGGCGTT TGCCTGGTTG TGGCCCGGCG
35 101 TGTTCCTCCA CGATTGTGG AATCCTGACG AACCTGCCGT CTATACCGCC
151 GTCGAAGCAC TGGCAGGCAG CCCCACCCCT TTGGTTGCC ATCTGTTCGG
201 TCAAATCGAT TTCGGCATA CGCCCGTGTA TCTTTGGGT GCGCCCGCGT
251 TCAAACATTT GCTGTGCGCG TGGGCTGCCG ACCCGTATGA TGCCGCACGC
40 301 TTTGCCGGCG TGTTCCTCG CGTTGTCCGA CTGACTTCCT GCGGCTTTGC
351 CGGTTTCAAC TTTTGGGCA GACACCACGG GCGCAGCGTC GTCCTGATTC
401 TCATCGGCTG TATCGGCTG ATTCCGACCG TACACTTTCT CAACCCCGCT
451 GCCCGCGCTT TTGCCCGCG CGGACTGGTG CTGCACGGTT ATTCTTTGGC
501 TCGCCGGCGC GTGATTGCCG CCTCTTTTCT GCTCGGTACG GGTGGGACGC
45 551 TGATGTCTGT GGCAGCAGCT TATCCGGCGG CATTGCCCCT GATGCTGCCC
601 CTGCCCCTGC TGATGTTTTT CCGTCCGTGG CAAAGCAGGC GTTTGATGTT
651 GACGGCAGTC GCCTCGCTTG CCTTTGCCCT GCCGCTTATG ACCGTTTACC
701 CGCTGCTCTT GGCAAAAACG CAGCCCGCGC TGTTCCGCGA ATGGCTCGAC
751 GATCACGTTT TCGGTACGTT CGGCGGCGTG CGGCACATTC AGACGGCATT
801 CAGTTTGTCT TACTATCTGA AAAACCTGCT TTGGTTTGCA TTGCCTGCGC
50 851 TGCCGCTGGC GGTTCGGACG GTTTGCCGCA CGCGCCTGTT TTCGACCGAC
901 TGGGGGATTT TGGCGCTCGT CTGGATGCTT GCCGTTTGG TGCTGCTTGC
951 CGTCAATCCG CAGCGTTTTT AGGATAACCT CGTCTGGCTG CTTCCGCCGC
1001 TTGCCCTGTT CGGCGCGGCG CAACTGGACA GCCTGAGACG CGGCGCGGCG
1051 GCGTTTGTCA ACTGTTTCGG CATTATGGCG TTCGGACTGT TTGCCGTGTT
55 1101 CCTGTGGACG GGCTTTTTCG CCATGAATTA CGGCTGGCCC GCCAAGCTTG
1151 CCGAACGCGC CGCTATTTTC AGCCCGTATT ATGTTCTTGA TATCGATCCC
1201 ATTCCGATGG CGGTTGCCGT ACTGTTTACA CCCTTGTTGG TGTGGGCGAT
1251 TACCCGCAAA AACATACGCG GCAGGCAGGC GGTACCAAC TGGGCGGCG
60 1301 GCGTTACCTT GACCTGGGCT TTGCTGATGA CGCTGTTCTT GCCGTGGCTG
1351 GACGCGGCGA AAAGCCACGC GCCCGTCGTC CGGAGTATGG AGGCATCGCT
1401 TTCCCGGAA TTAACCGGG AGCTTTCAGA CGGCATCGAG TGTATCGACA
1451 TAGGCGGCGG CGACCTACAC ACGCGGATTG TTTGGACGCA GTACGGCACA
1501 TTGCCGCACC GCGTCGGCGA TGTACAATGC CGCTACCGCA TCGTCCGCTT
1551 GCCCCAAAAC GCGGATGCGC CGCAAGGCTG GCAGACGGTC TGGCAGGGTG

```



-340-

1601 CGCGCCCGCG CAACAAAGAC AGTAAGTTCG CACTGATACG GAAAACCGGG  
 1651 GAAAATATAT TAAAAACAAC AGATTGA

This encodes a protein having amino acid sequence <SEQ ID 598>:

```

      1 MLTYTPPDAR PPAKTHEKPW LLLMAFAWL WPGVFSHDLW NPDEPAVYTA
5     51 VEALAGSPTP LVAHLFGQID FGIPPVYLWV AAFAKHLLSP WAADPYDAAR
      101 FAGVFFAVVG LTSCGFAGFN FLGRHHGRSV VLILIGCIGL IPTVHFLNPA
      151 AAFAAAGLV LHGYSLARRR VIAASFLLGT GWTLMSLAAA YPAAFALMLP
      201 LPVLMFFRPW QSRRMLTAV ASLAFALPLM TVYPLLAKT QPALFAQWLD
      251 DHVFGTFGGV RHIQTAFSLF YYLKNLLWEA LPALPLAVWT VCRTRLFSTD
10    301 WGILGVVWML AVLVLAVNP QRFQDNLVWL LPPLALFGAA QLDSLRRGAA
      351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAYF SPYYVPDIDP
      401 IPMAVAVLET PLWLWAITRK NIRGRQAVTN WAAGVTLTWA LLMTLFLPW
      451 DAAKSHAPVV RSMEASLSPE LKRELSDGIE CIDIGGGDLH TRIVWTQYGT
      501 LPHRVGDVQC RYRIVRLPQN ADAPQGWQTV WQGARPRNKD SKFALIRKTG
15    551 ENILKTTD*
  
```

ORF141a and ORF141-1 show 98.2% identity in 553 aa overlap:

```

      orf141a.pep MLTYTPPDARPPAKTHEKPWLLLMAFAWLWPGVFSHDLWNPDEPAVYTAVEALAGSPTP
      orf141-1    MLTYTPPDARPPAKTHEKPWLLLMAFAWLWPGVFSHDLWNPDEPAVYTAVEALAGSPTP
20    orf141a.pep LVAHLFGQIDFGIPPVYLWVAAAFKHLLSPWAADPYDAARFAGVFFAVVGLTSCGFAGFN
      orf141-1    LVAHLFGQIDFGIPPVYLWVAAAFKHLLSPWAADSYDAARFAGVFFAVVGLTSCGFAGFN
25    orf141a.pep FLGRHHGRSVVLILIGCIGLIPTVHFLNPAAAAFAAAGLVLHGYSLARRRVIAASFLLGT
      orf141-1    FLGRHHGRSVVLILIGCIGLIPVAHFLNPAAAAFAAAGLVLHGYSLARRRVIAASFLLGT
30    orf141a.pep GWTLMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRMLTAVASLAFALPLMTVYPLLAKT
      orf141-1    GWTLMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRMLTAVASLAFALPLMTVYPLLAKT
35    orf141a.pep QPALFAQWLDHVFVGTGGVVRHIQTAFSLFYLLKNLLWFALPALPLAVWTVCRTRLFSTD
      orf141-1    QPALFAQWLDYHVFVGTGGVVRHVQTAFSLFYLLKNLLWFALPALPLAVWTVCRTRLFSTD
40    orf141a.pep WGILGVVWMLAVLVLLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
      orf141-1    WGILGVVWMLAVLVLLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
45    orf141a.pep FGLFAVFLWTGFFAMNYGWPAKLAERAAYFSPYYVPDIDPIPMNAVAVLETPLWLWAITRK
      orf141-1    FGLFAVFLWTGFFAMNYGWPAKLAERAAYFSPYYVPDIDPIPMNAVAVLETPLWLWAITRK
50    orf141a.pep NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAKSHAPVVRSMASLSPELKRELSDGIE
      orf141-1    NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAKSHAPVVRSMASLSPELKRELSDGIE
55    orf141a.pep CIDIGGGDLHTRIVWTQYGTLPHRVGDVQCRIYRIVRLPQNADAPQGWQTVWQGARPRNKD
      orf141-1    CIDIGGGDLHTRIVWTQYGTLPHRVGDVQCRIYRIVRLPQNADAPQGWQTVWQGARPRNKD
      orf141a.pep SKFALIRKTGENI
      orf141-1    SKFALIRKIGENI
  
```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF141 shows 95% identity over a 140aa overlap with a predicted ORF (ORF141ng) from *N.gonorrhoeae*:

```

60    orf141.pep          DFGISPVYLWVAAAFKHLLSPWAADSYDVA      30
      orf141ng          WNPAEPAVYTAVEALAGSPTPLVAHLFGQIDFGIPPVYLWVAAAFKHLLSPWAADPYDAA 126
  
```

	orf141.pep	RFAGVFFAVIGLTS CGFAGFNFLGRHHGRXVVLILIGCIGLIPVAHFLNPAAAAFAAAGL	90
	orf141.ng	RFAGVFFAVIGLTS CGFAGFNFLGRHHGRSVVLIHIGCIGLIPVAHFFNPAAAAFAAAGL	186
5	orf141.pep	VLHGYSLARRRVIAAS FLLGTGWT LMSLAAAYPAA FALMLPLPVLMMFFRP	140
	orf141.ng	VLHGYSLARRRVIAAS FLLGTGWT LMSLAAAYPAA FALMLPLPVLMMFFRPQSRRLMLTA	246

An ORF141ng nucleotide sequence <SEQ ID 599> was predicted to encode a protein having amino acid sequence <SEQ ID 600>:

10	1	MPSEAVSARP	LCEYLLHLAI	RPFLTLMLT	YTPPDARPPA	KTHEKPWLLL
	51	LMAFAWLWPG	VFSDHLWNPA	EPAVYTAVEA	LAGSPTPLVA	HLFGQTDGFI
	101	PPVYLWVAAA	FKHLLSPWAA	HPYDAAREFAG	VEFAVIGLTS	CGFAGFNFLG
	151	RHHGRSVVLI	HIGCIGLIPV	AHFFNPAAAA	FAAAGLVHLG	YSLARRRVIA
	201	ASFLLTGTGWT	LMSLAAAYPA	AFALMLPLPV	LMFFRPQSR	RLMLTAVASL
15	251	AFALPLMTVY	PLLLAKTQPA	LFAQWLNYHV	FGTFGGVRHI	QRAFSLFHYL
	301	KNLLWFAPPG	LPLAVWTVCR	TRLFSTDWGI	LGIVWMLAVL	VLLAFNPQRF
	351	QDNLVWLLPP	LALFGAAQLD	SLRRGAAAFV	NWFGIMAFGL	FAVFLWTGFF
	401	AMNYGWPAKL	AERAAYFSFY	YVPDIDPIPM	AVAVLETPW	LWATRKNIR
	451	GRQAVTNWAA	GVTLTWALLM	TLFLPWLDA	KSHAPVVRSM	EASFSPELKR
20	501	ELSDGIECIG	IGGGDLHTRI	VWTQYGTLP	RVGDVRCRYR	IVRLPQNADA
	551	PQGWQTVWQG	ARPRNKDSKF	ALIRKIGENI	LKTTD*	

Further work revealed the following gonococcal DNA sequence <SEQ ID 601>:

	1	ATGCTGACCT	ATACCCCGCC	CGATGCCCGC	CCGCCCGCCA	AAACCCACGA
	51	AAAACCGTGG	CTGCTGCTGT	TGATGGCGTT	TGCCTGGCTG	TGGCCCGCGG
25	101	TGTTTTCCCA	CGATTGTGG	AATCCTGCCG	AACCTGCCGT	CTATACCGCC
	151	GTCGAAGCAC	TGGCAGGCAG	CCCCACCCCG	TTGGTTGCCC	ATCTGTTCCG
	201	TCAAACCGAT	TTCGGCATA	CGCCCGTGTA	TCTTTGGGTT	GCCGCCGCAT
	251	TCAAACATTT	GCTGTCGCCG	TGGGCAGCCG	ACCCGTATGA	TGCCGCACGC
	301	TTTGCAGGCG	TATTTTTTGC	CGTTATCGGA	CTGACTTCTT	GCGGCTTTGC
30	351	CGGTTTCAAC	TTTTTGGGCA	GACACCACGG	GCGCAGCGTT	GTTTTAATCC
	401	ATATCGGCTG	TATCGGGCTG	ATTCCGGTTG	CCCATTTCCT	CAATCCcgcc
	451	gcgcgcgcct	tGCCGCCCGC	CGGACTGGTG	CTGCacggct	actcgctgGC
	501	ACGCCGGCGC	GTGAttgccg	cctctTtccT	GCTCGGTACG	GGTTGGACGT
	551	TGATGTCGCT	GGCGGCAGCT	TATCCGGCGG	CGTTTGCCTG	GATGCTGCCC
35	601	CTGCCCGTGC	TGATGTTTTT	CCGTCCGTGG	CAAAGCAGGC	GTTTGATGTT
	651	GACGGCAGTC	GCCTCGCTTG	CCTTTGCCCT	GCCGCTTATG	ACCGTTTACC
	701	CGTGCTCtt	gGCAAAACG	CAGCCCGCGC	TGTTTGCACA	ATGGCTCAAC
	751	TATCACGTTT	TCGGTACGTT	cggcgGCGTG	CGGCAcattC	AGAggGCatT
	801	Cagtttgttt	cactatctgA	AAaatctgct	ttggttcgca	ccgcccgggC
40	851	TGCCGCTGGC	GTTTGGACG	GTTTGGCCGA	CACGCCTGTT	TTCGACCGAC
	901	TGGGGGATTT	TGGGCATTGT	CTGGATGCTT	GCCGTTTTGG	TGCTGCTCGC
	951	CTTTAATCCG	CAGCGTTTTT	AAGACAACCT	CGTCTGGCTG	CTGCCGCGCG
	1001	TTGCCCTGTT	CGGCGCGCGC	CAACTGGACA	GCCTGAGGCG	CGGCGCGGCG
	1051	GCTTTTGTC	ACTGGTTCGG	CATTATGGCG	TTCGGGCTGT	TGCGCGTGT
45	1101	CCTGTGGACG	GGCTTTTTTC	CCATGAATTA	CGGCTGGCCC	GCCAAGCTTG
	1151	CCGAACGCGC	CGCCTACTTC	AGCCCGTATT	ACGTTCCCGA	CATCGATCCC
	1201	ATTCCGATGG	CGGTTGCCGT	ACTGTTTACA	CCCTTGTTGG	TGTGGGCGAT
	1251	TACCCGGA	AACATACGCG	GCAGGCAGGC	GGTTACCAAC	TGGGCGGCAG
	1301	GCGTTACCTT	GACCTGGGCT	TTGCTGATGA	CGCTGTTCTT	GCCGTGGCTG
50	1351	GACGCGGCGA	AAAGCCACGC	GCCCGTCGTC	CGGAGTATGG	AGGCATCGTT
	1401	TTCCCCGGAA	TTAAACCGGG	AGCTTTCAGA	CGGCATCGAG	TGTATCGGCA
	1451	TAGGCGGCGG	CGACCTGCAC	ACGCGGATTG	TTTGGACGCA	GTACGGCACA
	1501	TTGCCGCACC	GCCTCGGCGA	TGTCCGTTGC	CGCTACCGTA	TCGTCCGCTT
	1551	GCCCCAAAC	GCGGATCGCG	CGCAAGGCTG	GCAGACGGTC	TGGCAGGGTG
55	1601	CGCGCCCGCG	CAACAAAGAC	AGTAAGTTTG	CACTGATACG	GAAAATCGGG
	1651	GAAAATATAT	TAAAAACAAC	AGATTGA		

This corresponds to the amino acid sequence <SEQ ID 602; ORF141ng-1>:

	1	MLTYTPPDAR	PPAKTHEKPW	LLLLMAFAWL	WPGVFSDHLW	NPAEPAVYTA
	51	VEALAGSPTP	LVAHLFGQTD	FGIPPVYLWV	AAAFKHLLSP	WAADPYDAAR
60	101	FAGVFFAVIG	LTSCGFAGFN	FLGRHHGRSV	VLIHIGCIGL	IPVAHFLNPA
	151	AAAFAAAGLV	LHGYSLARRR	VIAASFLLGT	GWTLMSLAAA	YPAAFALMLP
	201	LPVLMFFRPW	QSRRLMLTAV	ASLAFALPLM	TVYPLLLAKT	QPALFAQWLN
	251	YHVFGTFGGV	RHIQRAFSLF	HYLKNLLWFA	PPGLPLAVWT	VCRTRLFSTD
	301	WGILGIVWML	AVLVLLAFNP	QRFDNLVWL	LPPLALFGAA	QLDSLRRGAA

351 AFDVWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAYF SPYYVPDIDP  
 401 IPMAVAVLFT PLWLWAIKTRK NIRGRQAVTN WAAGVTLTWA LLMTLFLPLWL  
 451 DAAKSHAPVV RSMEASFSPK LKRELSKGIE CIGIGGGDLH TRIVWTQYGT  
 501 LPHRVGDVRC RYRIVRLPQN ADAPOGWQTV WQGARPRNKD SKFALIRKIG  
 551 ENILKTTD\*

ORF141ng-1 and ORF141-1 show 97.5% identity in 553 aa overlap:

```

10 orf141ng-1.pep MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPAPAVYTAVEALAGSPTP
    orf141-1      MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPDEPAVYTAVEALAGSPTP

15 orf141ng-1.pep LVAHLFGQTDGFI PPVYLWVAAAFKHLSPWAADPYDAARFAGVFFAVIGLTSCGFAGFN
    orf141-1      LVAHLFGQTDGFI PPVYLWVAAAFKHLSPWAADSYDAARFAGVFFAVIGLTSCGFAGFN

20 orf141ng-1.pep FLGRHHGRSVVLIIHIGCIGLIPVAHFLNPAAAAFAAGLVLHGYSLARRRVIAASFLGTT
    orf141-1      FLGRHHGRSVVLIIHIGCIGLIPVAHFLNPAAAAFAAGLVLHGYSLARRRVIAASFLGTT

25 orf141ng-1.pep GWTLMSLAAAYPAAFALMLPLPVLMEFRPWQSRRLMLTAVASLAFALPLMTVYPLLAKT
    orf141-1      GWTLMSLAAAYPAAFALMLPLPVLMEFRPWQSRRLMLTAVASLAFALPLMTVYPLLAKT

30 orf141ng-1.pep QPALFAQWLNHYVFGTGGVVRHIIQAFSLFHYLKNLLWFAPPGLPLAVWTVCRTRLFSTD
    orf141-1      QPALFAQWLDYHVFSGTGGVVRHVQTAFSLFYLYLKNLLWFALPALPLAVWTVCRTRLFSTD

35 orf141ng-1.pep WGILGIVWMLAVLVLLAFNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
    orf141-1      WGILGVVWMLAVLVLLAFNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA

40 orf141ng-1.pep FGLFAVFLWTGFFAMNYGWP AKLAERAAYFSPYYVPDIDPIPMVAVLFTPLWLWAIKTRK
    orf141-1      FGLFAVFLWTGFFAMNYGWP AKLAERAAYFSPYYVPDIDPIPMVAVLFTPLWLWAIKTRK

45 orf141ng-1.pep NIRGRQAVTNWAAGVTLTWALLMTLFLPLWLDAAKSHAPVVRSMESFSPKRELSKGIE
    orf141-1      NIRGRQAVTNWAAGVTLTWALLMTLFLPLWLDAAKSHAPVVRSMESLSPKRELSKGIE

50 orf141ng-1.pep CIGIGGGDLHTRIVWTQYGTLP HRVGDVRCRYRIVRLPQNADAPQGWQTVWQGARPRNKD
    orf141-1      CIGIGGGDLHTRIVWTQYGTLP HRVGDVQCRYRIVLLPQNADAPQGWQTVWQGARPRNKD

55 orf141ng-1.pep SKFALIRKIGENILKTTDX
    orf141-1      SKFALIRKIGENIX

```

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 72

50 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 603>:

```

1  ..CAATCCGCCA AATGGTTATC GGGCCAACT CTAGTCGGCA CAGCAATTGG
51 GATACGCGGG CAGATAAGC TTGGCGGCAA CCTGCATTAC GATATATTTA
101 CCGGCCGCGC ATTGAAAAG CCCGAATTTT TCCAATCAAG GAAATGGGCA
151 AGCGGTTTTC AGGTAGGCTA TACGTTTAA

```

55 This corresponds to the amino acid sequence <SEQ ID 604; ORF142>:

```

1  ..QSAKWLSGQT LVGTAIGIRG QIKLGGNLHY DIFTGRALKK PEFFQSRKWA
51 SGFQVGYTF*

```

Further work revealed the complete nucleotide sequence <SEQ ID 605>:

1 ATGGATAATT CGGGTAGTGA GGCGACAGGA AAATACCAAG GAAATATCAC  
 51 TTTCTCTGCC GACAATCCTT TGGGACTGAG TGATATGTTT TATGTAAATT  
 101 ATGGACGTTT GATTGGCGGT ACGCCCGATG AGGAAAAGTT TGACGGCCAT  
 151 CGCAAAGAAG GCGGATCAAA CAATTACGCC GTACATTATT CAGCCCCCTT  
 201 CGGTAAATGG ACATGGGCAT TCAATCACAA TGGCTACCGT TACCATCAGG  
 251 CAGTTTCCGG ATTATCGGAA GTCTATGACT ATAATGGAAA AAGTTACAAT  
 301 ACTGATTTTC GCTTCAACCG CCTGTTGTAT CGTGATGCCA AACGCAAAAC  
 351 CTATCTCAGT GTAAACTGT GGACGAGGGA AACAAAAAGT TACATTGATG  
 401 ATGCCGAAC TACTGTACAA CGGCGTAAAA CTGCGGGTTG GTTGGCAGAA  
 451 CTTTCCACCA AAGAATATAT CGGTGCGAGT ACGGCAGATT TTAAGTTGAA  
 501 ATATAAACGC GGCACCGGCA TGAAAGATGC TCTGCGCGCG CCTGAAGAAG  
 551 CTTTGGCGCA AGGCACGTCA CGTATGAAAA TTTGGACGGC ATCGGCTGAT  
 601 GTAAATACTC CTTTTCAAAT CGGTAAACAG CTATTGCGCT ATGACACATC  
 651 CGTTCATGCA CAATGGAAAC AAACCCCGCT AACATCGCAA GACAACTGG  
 701 CTATCGGCGG ACACCACACC GTACGTGGCT TCGACGGTGA AATGAGTTTG  
 751 TCTGCCGAGC GGGGATGGTA TTGGCGCAAC GATTGAGCT GGCAATTTAA  
 801 ACCAGGCCAT CAGCTTTATC TTGGGGCTGA TGTAGGACAT GTTTCAGGAC  
 851 AATCCGCCAA ATGGTTATCG GGCCAAACTC TAGTCGGCAC AGCAATTGGG  
 901 ATACGCGGGC AGATAAAGCT TGGCGGCAAC CTGCATTACG ATATATTTAC  
 951 CGGCCGCGCA TTGAAAAAGC CCGAATTTT CCAATCAAGG AAATGGGCAA  
 1001 GCGGTTTTCA GGTAGGCTAT ACGTTTTAA

This corresponds to the amino acid sequence <SEQ ID 606; ORF142-1>:

1 MDNSGSEATG KYQGNITFSA DNPLGLSDMF YVNYGRSIGG TPDEESFDGH  
 51 RKEGGSNNYA VHSAFFGKW TWFNHNHGYR YHQAUSGLSE VYDYNKSYN  
 101 TDFGFNRLLY RDAKRKTYLG VKLWMRETKS YIDDAELTVQ RRKTAGWLAE  
 151 LSHKEYIGRS TADFKLKYKR GTGMKDALRA PEEAFGEGTS RMKIWTASAD  
 201 VNTPFQIGKQ LFAYDTSVHA QWNKTPLTSQ DKLAIGGHHT VRGFDGEMSL  
 251 SAERGWYWRN DLSWQFKPGH QLYLGADVGH VSGQSAKWLS GQTLVGTAIG  
 301 IRGQIKLGGN LHYDIFTGRA LKKPEFFQSR KWASGFQVGY TF\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF142 shows 88.1% identity over a 59aa overlap with a predicted ORF (ORF142ng) from *N.gonorrhoeae*:

35	orf142.pep		QSAKWLSGQTLVGTAIGIRGQIKLGGNLHY	30
	orf142ng	RGWYWRNDLSWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLAGTAIGIRGQIKLGGNLHY	313	
	orf142.pep	DIFTGRALKKPEFFQSRKWASGFQVGYTF	59	
40	orf142ng	DIFTGRALKKPEYFQTKKWTGTFQVGYSF	342	

The complete length ORF142ng nucleotide sequence <SEQ ID 607> is:

1 ATGGATAATT CGGGTAGTGA GGCGACAGGA AAATACCAAG GAAATATCAC  
 51 TTTCTCTGCC GACAATCCTT TTGGACTGAG TGATATGTTT TATGTAAATT  
 101 ATGGACGTTT AATTGGCGGT ACGCCCGATG AGGAAAATTT TGACGGCCAT  
 151 CGCAAAGAAG GCGGATCAAA CAATTACGCC GTACATTATT CAGCCCCCTT  
 201 CGGTAAATGG ACATGGGCAT TCAATCACAA TGGCTACCGT TACCATCAGG  
 251 CAGTTTCCGG ATTATCGGAA GTCTATGACT ATAATGGAAA AAGTTACAAC  
 301 ACTGATTTTC GCTTCAACCG CCTGTTGTAT CGTGATGCCA AACGCAAAAC  
 351 CTATCTCAGT GTAAACTGT GGACGAGGGA AACAAAAAGT TACATTGATG  
 401 ATGCCGAAC TACTGTACAA CGGCGTAAAA CCACAGGTTG GTTGGCAGAA  
 451 CTTTCCACCA AAGGATATAT CGGTGCGAGT ACGGCAGATT TTAAGTTGAA  
 501 ATATAAACAC GGCACCGGCA TGAAAGATGC TCTGCGCGCG CCTGAAGAAG  
 551 CTTTGGCGCA AGGCACGTCA CGTATGAAAA TTTGGACGGC ATCGGCTGAT  
 601 GTAAATACTC CTTTTCAAAT CGGTAAACAG CTATTGCGCT ATGACACATC  
 651 CGTTCATGCA CAATGGAAAC AAACCCCGCT AACATCGCAA GACAACTGG  
 701 CTATCGGCGG ACACCACACC GTACGTGGCT TCGACGGTGA AATGAGTTTG  
 751 CTTGCCGAGC GGGGATGGTA TTGGCGCAAC GATTGAGCT GGCAATTTAA  
 801 ACCAGGCCAT CAGCTTTATC TTGGGGCTGA TGTAGGACAT GTTTCAGGAC  
 851 AATCCGCCAA ATGGTTATCG GGCCAAACTC TAGCCGGCAC AGCAATTGGG  
 901 ATACGCGGGC AGATAAAGCT TGGCGGCAAC CTGCATTACG ATATATTTAC  
 951 CGGCCGTGCA TTGAAAAAGC CCGAATATTT TCAGACGAAG AAATGGGTAA

1001 CGGGGTTTCA GGTGGGTAT TCGTTTTGA

This encodes a protein having amino acid sequence <SEQ ID 608>:

```

      1  MDNSGSEATG KYQGNITFSA DNPFGLSDMF YVNYGRSIGG TPDEENFDGH
      5  51  RKEGGSNNYA VHYSAPFGKW TWAENHNGYR YHQAUSGLSE VYDYNKSYN
      10  101 TDFGFNRLLY RDAKRKTYLS VKLWTRETKS YIDDAELTVQ RRKTGWLAE
      15  151 LSHKGYIGRS TADFKLKYKH GTGMKDALRA PEEAFGEGTS RMKIWTASAD
      20  201 VNTFFQIGKQ LFAYDTSVHA QWNKTPLTSQ DKLAIGGHHT VRGFDGEMSL
      25  251 PAERGWIWRN DLSWQFKPGH QLYLGADVGH VSGQSAKWLS GQTLAGTAIG
      30  301 IRGQIKLGGN LHYDIFTGRA LKKPEYFQTK KWTGTFQVGY SF*
  
```

10 The underlined sequence (aromatic-Xaa-aromatic amino acid motif) is usually found at the C-terminal end of outer membrane proteins.

ORF142ng and ORF142-1 show 95.6% identity over 342aa overlap:

```

      orf142-1.pep MDNSGSEATGKYQGNITFSADNPLGLSDMFYVNYGRSIGGTPDEESFDGHRKEGGSNNYA
      15  orf142ng-1   MDNSGSEATGKYQGNITFSADNPFGLSDMFYVNYGRSIGGTPDEENFDGHRKEGGSNNYA

      orf142-1.pep VHYSAPFGKWTWAFNHNGYRYHQAVSGLSEVYDYNKSYNTDFGFNRLLYRDAKRKTYLG
      20  orf142ng-1   VHYSAPFGKWTWAFNHNGYRYHQAVSGLSEVYDYNKSYNTDFGFNRLLYRDAKRKTYLS

      orf142-1.pep VKLWMRETKSYIDDAELTVQRRKTAGWLAELSHKEYIGRSTADFKLKYKRGTMKDALRA
      25  orf142ng-1   VKLWTRETKSYIDDAELTVQRRKTGWLAELSHKGYIGRSTADFKLKYKHGTGMKDALRA

      orf142-1.pep PEEAFGEGTSRMKIWTASADVNTFFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHT
      30  orf142ng-1   PEEAFGEGTSRMKIWTASADVNTFFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHT

      orf142-1.pep VRGFDGEMSLPAERGWIWRNDLSWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLVGTAG
      35  orf142ng-1   VRGFDGEMSLPAERGWIWRNDLSWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLAGTAIG

      orf142-1.pep IRGQIKLGGNLHYDIFTGRALKKPEFFQSRKWASGFQVGYTF
      40  orf142ng-1   IRGQIKLGGNLHYDIFTGRALKKPEYFQTKKWTGTFQVGYSF
  
```

In addition, ORF142ng is homologous to the HecB protein of *E. chrysanthemi*:

```

gi|1772622 (L39897) HecB [Erwinia chrysanthemi] Length = 558
Score = 119 bits (295), Expect = 3e-26
Identities = 88/346 (25%), Positives = 151/346 (43%), Gaps = 22/346 (6%)

40  Query: 2  DMSGSEATGKYQGNITFSADNPFGLSDMFYVNYGRSIGGTPDEENFDGHRKEGGSNNYAV 61
      DMSG ++TG+ Q N + + DN FGL+D ++++ G S + + D + G
      Sbjct: 230 DMSGQKSTGEEQLNGSLALDNVFLADQWFISAGHS---SRFATSHDAESLQAG----- 280

45  Query: 62 HYSAPFGKWTWAFNHNGYRYHQAVSGLSEVYDYNKSYNTDFGFNRLLYRDAKRKTYLSV 121
      +S P+G W +N++ RY + G S F +R+++RD KT ++
      Sbjct: 281 -FSMPYGYWNLGYNYSQSRYNRTFINRDFPWHSTGSDTHRFSLSRVVRDGTMTKTAIAG 339

50  Query: 122 KLWTRETKSYIDDAELTVQRRKTGWLAELSHKGYIGRSTADFKLKYKHGTGMKDALARAP 181
      R +Y++ + L RK + ++H + A F Y G +
      Sbjct: 340 TFSQRTGNNYLNGSLLPSSSRKLSSVSLGVNHSQKLWGGGLATFNPTYNRGVRWLGSSETDT 399

55  Query: 182 EEAFFGEGTSRMKIWTASADVNTFFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHTV 241
      +++ E + WT SA P Y S++ Q++ L ++L +GG ++
      Sbjct: 400 DKSADPRAEFNKWTLASYYHPV---TDSITYLGSLYGQYSARALYGSEQLTLGESSI 456

60  Query: 242 RGFDGEMSLPAERGWIWRNDLSWQFKP----GHQLYLGA-DVGHVSGQSAKWLSGQTLAG 296
      RGF E RG YWRN+L+WQ G+ ++ A D GH+ + +L G
      Sbjct: 457 RGF-REQYTSNGRGAYWRNELNWQAWQLPVLGNVTFMAAVDGGHLYNHKQDNSTAASLWG 515

      Query: 297 TAIGIRGQIKLGGNLHYDIFTGRALKKPEYFQTKKWTGTFQVGYSF 342
      A+G+ + L + G + P + Q V G++VG SF
  
```

### 5 Example 73

10

1	ATGCGGACGA	AATGGTCAGC	AGTGAGAAGC	TGCTTACTTG	GgCGGACACC
51	GCCGACATCG	ATACCGCTTT	GAACCTGTTG	TACCGTTTGC	AAAAACTTCGA
101	ATTCTCTAT	GCGGATGAA	ACGGTCATTC	AGACGGCATC	AAATTTGwCGG
151	ACAGCAATT	GGCTTGCTG	ATGGAACAAT	TGTCCGGCAG	CGGTAAAGCG
201	TTATTGGTCG	ATCGGAACGG	TCTGTATCTT	CCCAACGCCA	ATTTCCATCA
251	TGAGGCGGCG	GAAGAGTTGG	GTTGTTTGGC	GGCAGAAGTC	GCACAGATGG
301	AAAAGAAATA	CCGGCTGCTG	ATTAAGAACA	AC..	

```

15      1  MRTKWSAVRS  CTWADTADID  TALNLLYRLQ  KLEFLYGDEN  GHSDGINLXD
      51  EQLPLLMEQL  SGSGKALLVD  RNgLYLANAN  FHHEAAEELG  LLAAEVAQME
     101  KKYRLLIKNN  ..

```

20	1	ATGGAATCAA	CACTTTC	ACT	ACAAGCAAAT	TTATATCCCC	GCCTGACTCC
	51	TGCCGGTGCA	TTTTATGCCG	TATCCAGCGA	TGCCCCCAGT	GCCGGTAAAA	
	101	CTTTGTTGCA	CAGCCTGTTG	AAAGCAGATG	CGGACGAAAT	GGTCAGCAGT	
	151	GAGAAGCTGC	TTACTTGGGC	GGACACCGCC	GACATCGATA	CCGCTTTGAA	
	201	CCTGTTGTAC	CGTTTGCAAA	AACTCGAATT	CCTCTATGGC	GATGAAAACG	
25	251	GTCATTGAGA	CGGCATCAAT	TTGTCGGACG	AGCAATTGCC	GTTGCTGATG	
	301	GAACAATTGT	CCGGCAGCGG	TAAGGCGTTA	TTGGTCGATC	GAACCGGTCT	
	351	GTATCTTGCC	AACGCCAATT	TCCATCATGA	GGCGGGCGAA	GAGTTGGGGT	
	401	TGTTGGCGGG	AGAAAGTCGA	CAGATGGAAA	AGAAATACCG	GCTGCTGATT	
	451	AAGAACAACC	TGTATATCAA	CAATAACGCT	TGGGGCGGTT	GCATCTCTTC	
30	501	CGGTGAGAG	GAAATGACAT	TTTTCCCAT	GTATATCGGT	TCAACCAAT	
	551	TTATTTTGGT	TATCGGCGGC	ATTCGCGATT	TGGGCAAGAA	GGCATTGTGT	
	601	ACTTTGTGTA	GGATTTTATA	CCGCCGTTAC	AGCAACCGCG	TGTAA	

35

1	MESTLSLQAN	LYPRLTPAGA	FYAVSSDAPS	AGKTLHSL	KADADEMVSS
51	EKLLTWADTA	DIDTALNLLY	RLQKLEFLYG	DENGHSDDGIN	LSDEQLPLLM
101	EQLSGSGKAL	LVDRCPLYLA	NANFHHAEAE	ELGLLAAEVA	QMEKKYRLLI
151	KNNLYINNNA	WGVCNPSGQS	<u>ELTFFFPLYIG</u>	<u>STKFILVIGG</u>	IPDLGKEAFV
201	TLVRILYRRY	SNRV*			

### Homology with a predicted ORF from *N.meningitidis* (strain A)

```

                                     10      20      30
orif143.pep                      MRTKWSAVRSCTWADTADIDTALNLLYRLQKLEFL
                                     | : : ||| ||||| ||||| |||||
45 orf143a    GAFYAVSSDXPSAGKTLHSLLKADADEMVSSEKLLTWAXTADIDTALNLLYRLQKLEFL
               20       30       40       50       60       70
                               40       50       60       70       80       90
orif143.pep    YGDENGHSDGINLXDEQLPLLMEQLSGSGKALLVDRNGLYLANANFHEAAEELGLLAAE
50              ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

orf143a      YGDENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLANANFHHAAEELGLLAAE
              80      90      100      110      120      130
5  orf143.pep      100      110
              VAQMEKKYRLLIKNN
              |||||
orf143a      VAQMEKKYRLXIKNNLYINNNNAWGVCPSGQSELTFFPLYIGSTKFILVIGGIPDLGKEA
              140      150      160      170      180      190

```

The complete length ORF143a nucleotide sequence <SEQ ID 613> is:

```

10      1  ATGGAATCAA CANTTTCACT ACAAGCAAAT TTATATCNCC GCCTGACTCC
      51  TGCCGGTGCA TTTTATGCCG TATCCAGCGA TGNCCCCAGT GCCGGTAAAA
     101  CTTTGTGTCG CAGCCTGTTG AAAGCGGATG CGGACGAAAT GGTNAGCAGT
     151  GAGAAGCTGC TTACCTGGGC GGANACCGCC GACATCGATA CCGCTTTGAA
     201  CCTGTTGTAC CGTTTGCAA AACTCGAATT CCTCTATGGC GATGAAAACG
15      251  GTCAATCAGA CGGCATCAAT TTGTCGGACG AGCAATTGCC GTTGCTGATG
     301  GAACAATTGT CCGGCAGCGG TAAGGCGTTA TTGGTCGATC GGAACGGTCT
     351  GTATCTTGCC AACGCCAATT TCCATCATGA GGCGGCGGAA GAGTTGGGGT
     401  TGTGGCGGC AGAAGTCGCA CAGATGGAAA AGAAATACCG GCTGCNNATT
     451  AAGAACAACC TGTATATCAA CAATAACGCT TGGGGCGTTT GCGATCCTTC
20      501  CCGTCAGAGC GAATTGACAT TTTTCCCAT GTATATCGGT TCAACCAAAT
     551  TTATTTGGT TATCGGCGGC ATTCCGATT TGGGCAAAGA GGCATTGTGT
     601  ACTTTGGTAA GGATNTTATA CCNCCNGTTA CAGCAACCGC GTGTAAACT
     651  TGGGAGAGAG GANGGGTTAT GCAGCAATTA TTGA

```

This encodes a protein having amino acid sequence <SEQ ID 614>:

```

25      1  MESTXSLQAN LYXRLTPAGA FYAVSSDXPS AGKTLHSLK KADADEMVSS
     51  EKLLTWAXTA DDTALNLLY RLQKLEFLYG DENGHS DGIN LSDEQLPLLM
     101  EQLSGSGKAL LVDRNGLYLA NANFHHEAAE ELGLLAAEVA QMEKKYRLXI
     151  KNNLYINNNA WGVCDPSGQS ELTFFPLYIG STKFILVIGG IPDLGKEAFV
     201  TLVRXLYXXL QQPRVKLGRE XGLCSNY*

```

30 ORF143a and ORF143-1 show 97.1% identity in 207 aa overlap:

```

      orf143a.pep      MESTXSLQANLYXRLTPAGAFYAVSSDXPSAGKTLHSLKADADEMVSSEKLLTWAXTA
      orf143-1         MESTLSLQANLYPRLTGAFYAVSSDAPSAGKTLHSLKADADEMVSSEKLLTWADTA
35      orf143a.pep      DIDTALNLLYRLQKLEFLYG DENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA
      orf143-1         DIDTALNLLYRLQKLEFLYG DENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA
40      orf143a.pep      NANFHHEAAEELGLLAAEVAQMEKKYRLXIKNNLYINNNNAWGVCPSGQSELTFFPLYIG
      orf143-1         NANFHHEAAEELGLLAAEVAQMEKKYRLLIKNNLYINNNNAWGVCPSGQSELTFFPLYIG
      orf143a.pep      STKFILVIGGIPDLGKEAFVTLVRXLY
45      orf143-1         STKFILVIGGIPDLGKEAFVTLVRILY

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF143 shows 95.5% identity over a 110aa overlap with a predicted ORF (ORF143ng) from *N.gonorrhoeae*:

```

50      orf143.pep      MRTKWSAVRSCTWADTADIDTALNLLYRLQKLEFLYG DENGHS DGINLXDEQLPLLMEQL      60
      orf143ng         MRTKWSAVRSCSRADTADIDTALNLLYRLQKLEFLYG DENGHS DGINLSDEQLPLLMEQL      60
55      orf143.pep      SGSGKALLVDRNGLYLANANFHHAAEELGLLAAEVAQMEKKYRLLIKNN      110
      orf143ng         SGSGKALLVDRNGLYLANANFHHEAAEELGLLAAEVAQMEKKYRLLIIRNNLYINNNNAWGV      120

```

An ORF143ng nucleotide sequence <SEQ ID 615> was predicted to encode a protein having amino acid sequence <SEQ ID 616>:

```

1  MRTKWSAVRS  CSRADTADID  TALNLLYRLQ  KLEFLYGDEN  GHSDDGINLSD
51  EQPLLLMEQL  SGSGKALLVD  RNgLYLANAN  FHHESAEELG  LLAAEVAQME
101 KKYRLLRNN  LYINNNAWGV  CDPGQSELT  FFPLYIGSTK  FILVIAGIPD
151 LSKGGICYFG  KDFIPPLQQP  RVKLGTTGIM  RQLLSTIED  LNNSTSDIA
201 SAVISTDGLP  MATMLPSHLN  SDRVGAISAT  LLALGSRSVQ  ELACGELEQV
251 MIKKGSGYIL  LSQAGKDAVL  VLVAKETGRL  GLILLDAKRA  ARHIAEAT*

```

Further work revealed the following gonococcal DNA sequence <SEQ ID 617>:

	1	ATGGAATCAA	CAC TTTCACT	ACAAGCGAAT	TTATATCCCT	GCCTGACTCC
	51	TGCCGGTGCA	TTTTATGCCG	TATCCAGCGA	TGCCCCAGT	GCCGGTAAAA
10	101	CTTTGTTCGC	CAGCCTGTTG	AAAGCGGATG	CGGACGAAGT	GGTCAGCAGT
	151	GAGAAGCTGC	TCGCGGCGGA	CACCGCCGAC	ATCATGACCG	CTTTGAACCT
	201	GTTGTACCGT	TTGCAAAAAC	TCGAATTCCT	CTATGGCGAT	GAAAACGGTC
	251	ATTCAGACGG	CATCAATTTC	TCGGACGAGC	AAATGCCGTT	GCTGATGGAA
	301	CAATTGTCGG	GCAGCGGTAA	GGCATTATTG	GTCGATCGGA	ACGGTCTGTA
15	351	TCTTGCCAAC	GCCAATTTC	ATCATGAGTC	GGCGGAAGAG	TGTGGGTTGT
	401	TGGCGGCAGA	AGTCGCACAG	ATGGAAGAGA	AATACCGGCT	GCTGATTAGG
	451	AACAACCTGT	ATATCAACAA	TAACGCTTGG	GGCGTTTGCG	ATCCTTCCGG
	501	TCAGAGCGAA	TTGACATTTT	TCCCATTGTA	TATCGGTTCA	ACCAAATTTA
	551	TTTGTGTTAT	CGCCGGCATT	CCCGATTGTA	GCAAGAGGCG	ATTTGTTACT
20	601	TTGGTAAGGA	TTTTATACCG	CCGTTACAGC	AACCGCGTGT	AA

This corresponds to the amino acid sequence <SEQ ID 618; ORF143ng-1>:

25

1	MESTLSLQAN	LYPCLTPAGA	FYAVSSDAPS	AGKTLRLSL	KADADEVVSS
51	EKLLAADTAD	IDTALNLLYR	LQKLEFLYGD	ENGHSDGINL	SDEQLPILME
101	QLSGSGKALL	VDRNGPLYAN	ANFHESAE	LGLLAAEVAQ	MEKKYRLILR
151	NNLYINNNAW	GVCDPSGQSE	LTFFPLYIGS	TKFILVIAGI	PDLSEAFVLT
201	LVRILYRRYS	NRV*			

ORF143ng-1 and ORF143-1 show 95.8% identity in 214 aa overlap:

	orf143ng-1.pep	MESTLSLQANLYPCLTPAGAFYAVSSDAPSAGKTLRLSLKADAEVVSSEKLLA-ADTA	59
		:     :     :	
30	orf143-1	MESTLSLQANLYPRLTPAGAFYAVSSDAPSAGKTLHLKSLKADAEVVSSEKLLTWADTA	60
	orf143ng-1.pep	DIDTALNLLYRLQKLEFLYGDENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA	119
		:     :     :     :     :     :     :	
	orf143-1	DIDTALNLLYRLQKLEFLYGDENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA	120
35	orf143ng-1.pep	NANFHHESAEEGLLLAAEVAQMEKKYRLILRNLYINNNAWGCDPSGQSELTFFFLYIG	179
		:     :     :     :     :     :     :	
	orf143-1	NANFHHESAEEGLLLAAEVAQMEKKYRLILKNLYINNNAWGCDPSGQSELTFFFLYIG	180
40	orf143ng-1.pep	STKFILVIAGIPDLSKEAFVTLVRILYRRYSNRV	213
		:     :     :	
	orf143-1	STKFILVIGGIPDLGKEAFVTLVRILYRRYSNRV	214

Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 74

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 619>:

	1	ATGACCTTTT	TACAACGTTT	GCAAGGTTTG	GCAGACAATA	AAATCTGTGC
	51	GTTTGCATGG	TTCTGCGTCC	GCCGCTTTGA	TGAAGAACGC	GTACCGCAGr
50	101	CGGCGGCAAG	CATGACGTTT	ACGACGCTGC	TGGCATTCTG	CCCCGTGCTG
	151	ACCGTGATGG	TGGCGGTGCG	TTCGATTTTC	CCCGTGTTCC	ACCGGTGGTC
	201	GGATTCTGTT	GTCTCCTTCG	TCAACCAAAC	CATTGTGCCG	CA. GGCGCGG
	251	ACATGGTGTT	CGACTATATC	AATGCGTTCC	GCGAGCAGGC	GAACCGGCTG
	301	ACGGCAATCG	GCAGCGTGAT	GCTGGTCGTT	ACCTCGCTGA	TGCTGATTCT
55	351	GACGATAGAC	AATACGTTCA	ACCGCATCTG	G <u>a</u> CGGGTCAA	wTyCCAGCGT
	401	CCGTGGATG.	.			



This corresponds to the amino acid sequence <SEQ ID 620; ORF144>:

```

1  MTFLLRLQGL ADNKICAFW FVRRFDEER VPQAAASMTF TLLALVPVL
51  TVMVASIF PVFDRWDSF VSFVNQITVP XGADMVFDYI NAFREQANRL
101 TAIGSVMLVV TSLMLIRTID NTFNRIWRVX XQRPWM...

```

5 Further work revealed the complete nucleotide sequence <SEQ ID 621>:

```

1  ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
51  GTTTGCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCGCAGG
101 CGGCGGCAAG CATGACGTTT ACGACGCTGC TGGCACTCGT CCCCCTGCTG
151 ACCGTGATGG TGGCGGTCCG TTCGATTTTC CCCGTGTTCC ACCGCTGGTC
10  201 GGATTCGTTT GTCTCCTTCG TCAACCAAAC CATTTGTGCCG CAGGGCGCGG
251 ACATGGTGTT CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG
301 ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCTGA TGCTGATTCG
351 GACGATAGAC AATACGTTCA ACCGCATCTG GCGGGTCAAT TCCCAGCGTC
401 CGTGGATGAT GCAGTTTCTC GTCTATTGGG CTTTACTGAC GTTCGGGCCG
15  451 CTGTCTTTGG GCGTGGGCAT TTCCTTTATG GTCGGCTCGG TACAGGATGC
501 CGCGCTTGCC TCAGGTGCGC CGCAGTGGTC GGGCGCGTTG CGAACGCGCG
551 CGACGCTGAC CTTTCATGACG CTTTTGCTGT GGGGGCTGTA CCGCTTCGTG
601 CCAAACCGCT TCGTTCCCGC GCGGCAGGCG TTTGTGCGGG CTTTGGCAAC
651 AGCGTTTGTG CTGGAACCGC CGCGCTCCCT CTTCACTTGG TATATGGGCA
20  701 ATTTTCGACG CTACCGCTCG ATTTACGGCG CGTTTGCCGC CGTGCCGTTT
751 TTTCTGTGTG GGCTGAACCT GTTGTGACG CTGGTCTTGG GCGGCGCGGT
801 GCTGACTTCT TCACTCTCCT ACTGGCAGGG AGAAGCGTTC CGCAGGGGCT
851 TCGACTCGCG CGGACGTTT GACGACGTGT TGAATACTCT GCTGCTTCTG
901 GATGCGGCGC AAAAAGAAGG CAAAGCCTTG CCTGTTCAAG AGTTCAGACG
25  951 GCATATCAAT ATGGGCTACG ACGAGTTGGG CGAGCTTTTG GAAAAGCTGG
1001 CGCGGCACGG CTACATCTAT TCCGGCAGAC AGGGTTGGGT GTTGAAAACG
1051 GGGCGGATT CGATTGAGTT GAACGAACCT TTCAAGCTCT TCGTTTACCG
1101 TCCGTTGCCT GTGGAAAGGG ATCATGTGAA CCAAGCTGTC GATGCGGTAA
1151 TGACACCGTG TTTGCAGACT TTGAACATGA CGCTGGCAGA GTTTGACGCT
30  1201 CAGGCGAAAA AACGCGAGTA G

```

This corresponds to the amino acid sequence <SEQ ID 622; ORF144-1>:

```

1  MTFLLRLQGL ADNKICAFW FVRRFDEER VPQAAASMTF TLLALVPVL
51  TVMVASIF PVFDRWDSF VSFVNQITVP QGADMVFDYI NAFREQANRL
101 TAIGSVMLVV TSLMLIRTID NTFNRIWRVN SQRPWMMQFL VYWALLTFGP
35  151 LSLGVGISEM VGSVQDAALA SGAPQWSGAL RTAATLTMT LLLWGLYREV
201 PNRFPARQA FVGALATAFC LETARSLTW YMGNFDGYRS IYGAFAAVPF
251 FLLWLNLLWT LVLGGAVLTS SLSYWGGEAF RRGFDSRGRF DDVLKILLLL
301 DAAQKEGKAL PVQEFRRHIN MGYDELGELL EKLARHGYYI SGRQGWVLT
351 GADSIENEL FKLFPYRPLP VERDHVNQAV DAVMTPLQT LNMTLAEFDA
40  401 QAKKRQ*

```

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF144 shows 96.3% identity over a 136aa overlap with an ORF (ORF144a) from strain A of *N. meningitidis*:

```

45  orf144.pep 10 20 30 40 50 60
      MTFLLRLQGLADNKICAFWFVRRFDEERVVPQAAASMTFTLLALVPVLTVMVASIF
      orf144a 10 20 30 40 50 60
      MTFLLRLQGLADNKICAFWFVRRFDEERVVPQAAASMTFTLLALVPVLTVMVASIF
50  orf144.pep 70 80 90 100 110 120
      PVFDRWSDSFVSFVNQITVPXGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID
      orf144a 70 80 90 100 110 120
      PVFDRWSDSFVSFVNQITVPQGADMVFDYINAFREQANRLTAIGSVMLVVTXMLIRTID
55  orf144.pep 130
      NTFNRIWRVXXQRPWM
      orf144a 130
      NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISEFXVGSVQDAALASGAPQWSGAL
60

```

130 140 150 160 170 180

The complete length ORF144a nucleotide sequence <SEQ ID 623> is:

```

1  ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
5  51  GTTTGCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCGCAGG
101 CGGCGGCAAG CATGACGTTT ACGACACTGC TGGCACTCGT CCCCGTGTCTG
151 ACCGTGATGG TGGCGGTCCG TTCGATTTTC CCCGTGTTTC ACCGNTGGTC
201 GGATTCGTTT GTCTCCTTCG TCAACCAAAC CATTGTGCCG CAGGGCGCGG
251 ACATGGTNTT CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG
301 ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCNGA TGCTGATTCG
10 351 GACGATAGAC AATACGTTCA ACCGCATCTG GCGGGTCAAT TCCCAGCGTC
401 CGTGGATGAT GCAGTTTCTC GTCTATTGGG CTTTACTGAC GTTCGGGCCG
451 CTGTCTTTGG GCGTGGGCAT TTCCTTTATN GTCGGCTCGG TACAGGATGC
501 CGCGCTTGCC TCAGGTGCGC CGCAGTGGTC GGGCGCGTTG CGAACGGCGG
551 CGACGCTGAN CTTTCATGACG CTTTGTCTGT GGGGGCTGTA CCGCTNCGTG
15 601 CCAAACCGCT TCGTTCCTCG CCGGCANGCG TTTGTCTGGG CTTTGGCAAC
651 AGCGTTCTGT CTGGAACCGG CGCGTTCCTT CTTTACTTGG TATATGGGCA
701 ATTTTCGACG CTACCGCTCG ATTTACGGNG CGTTTGCCGC CGTGCCGTTT
751 TTTCTGTTGT GGCTGAACCT GTTGTGGACG CTGGTCTTGG GCGGCGCGGT
801 GCTGACTTCT TCACTCTCCT ACTGGCAGGG AGAAGCGTTC CGCAGGGNCT
20 851 TCGACTCGCG CGGACGGTTT GACGACGTGT TGAAAATCCT GCTGCTTCTG
901 GATGCGGCGC AAAAAGAAGG CNAAGCCTTG CCTGTTTCAGG AGTTCAGACG
951 GCATATCAAT ATGGGCTACG ACGAGTTGGG CGAGCTTTTG GAAAAGCTGG
1001 CGCGGCACGG CTACATCTAT TCCGGCAGAC AGGGTTGGGT GTTGAAAACG
1051 GGGGCGGATT CGATTGAGTT GAACGAATC TCAAGCTCT TCCTTTACCG
25 1101 TCCGTTGCC TGTGAAAGGG ATCATGTGAA CCAAGCTGTC GATGCGGTAA
1151 TGATGCCGTG TTTGCAGACT TTGAACATGA CGCTGGCAGA GTTTGACGCT
1201 CAGGCGAAAA AACAGCAGCA ATCTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 624>:

```

1  MTFLLRLQGL ADNKICAFW FVRRFDEER VPQAAASMTF TLLALVPVL
30 51  TVMVAVASIF PVFDRWDSF VSFVNQITVP QGADMVFDYI NAFREQANRL
101 TAIGSVMLV TSXMLIRTID NTFNRIWRVN SQRPMWQFL VYWALLTFGP
151 LSLGVGISFX VGSVQDAALA SGAPQWSGAL RTAATLXFMT LLLWGLYRXV
201 PNRFPARXA FVGALATAFC LETARSLFTW YMGNFDDYRS IYGAFAAVPF
251 FLLWLNLLWT LVLGGAVLTS SLSYWQGEAF RRXFDSRGRF DDVLKILLLL
35 301 DAAQKEGXAL PVQEFRRHIN MGYDELGELL EKLARHGYYI SGRQGWVLT
351 GADSIENEL FKLFEVYRPLP VERDHVNQAV DAVMMPCLOT LNMTLAEFDA
401 QAKKQQQS*

```

ORF144a and ORF144-1 show 97.8% identity in 406 aa overlap:

```

40 orf144a.pep MTFLLRLQGLADNKAFAFWFVRRFDEERVVPQAAASMTFTLLALVPVLTVMVAVASIF
orf144-1 MTFLLRLQGLADNKAFAFWFVRRFDEERVVPQAAASMTFTLLALVPVLTVMVAVASIF

orf144a.pep PVFDRWDSFVSFVNQITVPQGADMVFDYINAFREQANRLTAIGSVMLVVTTSXMLIRTID
45 orf144-1 PVFDRWDSFVSFVNQITVPQGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID

orf144a.pep NTFNRIWRVNSQRPMWQFLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL
orf144-1 NTFNRIWRVNSQRPMWQFLVYWALLTFGPLSLGVGISFMVGSVQDAALASGAPQWSGAL

50 orf144a.pep RTAATLXFMTLLLWGLYRXVPNRFPARXAFVGALATAFCLETARSLFTWYMGNFDDYRS
orf144-1 RTAATLTFMTLLLWGLYRXVPNRFPARQAFVGALATAFCLETARSLFTWYMGNFDDYRS

orf144a.pep IYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRXFDSRGRFDDVLKILLLL
55 orf144-1 IYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILLLL

orf144a.pep DAAQKEGXALPVQEFRRHINMGYDELGELLEKLARHGYYISGRQGWVLTGADSIENEL
60 orf144-1 DAAQKEGXALPVQEFRRHINMGYDELGELLEKLARHGYYISGRQGWVLTGADSIENEL

orf144a.pep FKLFEVYRPLPVERDHVNQAVDAVMMPCLOTLNMTLAEFDAQAKKQQQS 408
65 orf144-1 FKLFEVYRPLPVERDHVNQAVDAVMTPCLOTLNMTLAEFDAQAKKQQ 406

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF144 shows 91.2% identity over a 136aa overlap with a predicted ORF (ORF144ng) from *N.gonorrhoeae*:

5	orf144.pep	MTFLQRLQGLADNKICAFWVVRVFDEERVPOXAASMTFTTLLALVPVLTVMVAVASIF	60
	orf144ng	MTFLQCWQGSADNKICAFWVIRRFSEERVPOXAASMTFTTLLALVPVLTVMVAVASIF	60
	orf144.pep	PVFDRWSDSFVSFVNQTIQVPGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID	120
10	orf144ng	PVFDRWSDSFVSFVNQTIQVPGADMVFDYIDAFRDQANRLTAIGSVMLVVTSLMLIRTID	120
	orf144.pep	NTFNRIWRVXXQRPWM	136
15	orf144ng	NAFNRIWRVNTQRPWMQFLVYWALLTFGLSLGVGISFMVGSVQDSVLSSGAQQWADAL	180

The complete length ORF144ng nucleotide sequence <SEQ ID 625> is predicted to encode a protein having amino acid sequence <SEQ ID 626>:

	1	MTFLQCWQGS	ADNKICAFW	FVIRRFSEER	VPQAAASMTF	TTLALVPVL
	51	TVMVAVASIF	PVFDRWSDSF	VSFVNQTIQV	QGADMVFDYI	DAFRDQANRL
20	101	TAIGSVMLV	TSLMLIRTID	NAFNRIWRVN	TQRPWMMQFL	VYWALLTFGP
	151	LSLGVGISFM	VGSVQDSVLS	SGAQWADAL	KTAARLAEMT	LLLWGLYRFV
	201	PNRFVPAQQA	FVGALITAF	LETARFLFTW	YMGNFQGYRS	IYGAFVAVPF
	251	FLLWLNLLWT	LVLGGAVLTS	SLSYWQGEAF	RRGFDSRGRF	DDVLKILLLL
	301	DAAQKEGRTL	SVQEFRRHIN	MGYDELGELL	EKLARYGYIY	SGRQGWVLKT
25	351	GADSIELSEL	EKLARYGYIY	SGRQGWVLKT	LNMTLAEFDA	
	401	QAKKQQQS*				

Further work revealed the following gonococcal DNA sequence <SEQ ID 627>:

	1	ATGACCTTTT	TACAACGTTG	GCAAGGTTTG	GCGGACAATA	AAATCTGTGC
	51	ATTTCATGG	TTCGTCATCC	GCCGTTTCAG	TGAAGAGCGC	GTACCGCAGG
30	101	CAGCGGCGAG	CATGACGTTT	ACGACACTGC	TGGCACTCGT	CCCCGTAAGT
	151	ACCGTAATGG	TCGCGGTCGC	TTCGATTTTC	CCCGTGTTTC	ACCGCTGGTC
	201	GGATTCGTTT	GTCTCCTTCG	TCAACCAAAC	CATTGTGCGG	CAGGCGCGCG
	251	ATATGGTGTT	CGACTATATC	GACGCATTCC	GCGATCAGGC	AAACCGGCTG
	301	ACCGCCATCG	GCAGCGTGAT	GCTGGTCGTA	ACCTCGCTGA	TGCTGATTCG
35	351	GACGATAGAC	AATGCGTTCA	ACCGCATCTG	GCGGGTTAAC	ACGCAACGCC
	401	CCTGGATGAT	GCAGTTCTTC	GTTTATTGGG	CGTTGCTGAC	TTTCGGGCGT
	451	TTGCTTTTGG	GTGTGGGCAT	TTCCTTTATG	GTCGGGTCGG	TTCAAGACTC
	501	CGTACTCTCC	TCCGGAGCGC	AACAATGGGC	GGACGCGTTG	AAGACGGCGG
	551	CAAGGCTGGC	TTTCATGACG	CTTTTGCTGT	GGGGGCTGTA	CCGCTTCGTG
40	601	CCCAACCGCT	TCGTGCGCGC	CCGCGAGGCG	TTTGTGCGAG	CTTTGATTAC
	651	GGCATTCTGC	CTGGAGACGG	CACGTTTCCT	GTTCACTGGT	TATATGGGCA
	701	ATTTGACGG	CTACCGCTCG	ATTACGGCG	CATTGTGCGG	CGTGCCGTTT
	751	TTCTGCTGT	GGTAAACCT	GCTGTGGACG	CTGGTCTTGG	GCGGGGCGGT
	801	GCTGACTTCG	TCGCTGTCTT	ATTGGCAGGG	CGAGGCCTTC	CGCAGGGGAT
45	851	TCGACTCGCG	CGGACGGTTT	GACGACGTGT	TGAAAATCCT	GCTGCTTCTG
	901	GATGCGGCGC	AAAAGAAGG	CCGAACCTG	TCCGTTTCTG	AGTTTCTGAG
	951	GCATATCAAT	ATGGGTACG	ATGAATTGGG	CGAGCTTTTG	GAAAAGCTGG
	1001	CGCGGTACGG	CTATATCTAT	TCCGGCAGAC	AGGGCTGGGT	TTTGAAAACG
	1051	GGGGCGGATT	CGATTGAGTT	GAGCGAACTC	TTCAAGCTCT	TCGTGTACCG
50	1101	CCCGTTGCct	gtggaAAGGG	ATCATGTGAA	CCAAGCTGtc	gaTGCGGTAA
	1151	TGAcgccgtG	TTTGCAACT	TTGAACATGA	CGCTGGCGGA	GTTTGACGCT
	1201	CAGGcgAAAA	AACAGCAGCA	GTCTTGA		

This encodes a variant of ORF144ng, having the amino acid sequence <SEQ ID 628; ORF144ng-1>:

	1	MTFLQWQGL	ADNKICAFW	FVIRRFSEER	VPQAAASMTF	TTLALVPVL
55	51	TVMVAVASIF	PVFDRWSDSF	VSFVNQTIQV	QGADMVFDYI	DAFRDQANRL
	101	TAIGSVMLV	TSLMLIRTID	NAFNRIWRVN	TQRPWMMQFL	VYWALLTFGP
	151	LSLGVGISFM	VGSVQDSVLS	SGAQWADAL	KTAARLAEMT	LLLWGLYRFV
	201	PNRFVPAQQA	FVGALITAF	LETARFLFTW	YMGNFQGYRS	IYGAFVAVPF
	251	FLLWLNLLWT	LVLGGAVLTS	SLSYWQGEAF	RRGFDSRGRF	DDVLKILLLL
60	301	DAAQKEGRTL	SVQEFRRHIN	MGYDELGELL	EKLARYGYIY	SGRQGWVLKT

351 GADSIELSEL FKL FVYRPLP VERDHVNQAV DAVMTPCLQT LNM TLAEFDA  
401 QAKKQQQS\*

ORF144ng-1 and ORF144-1 show 94.1% identity in 406 aa overlap:

```

5  orf144ng-1.pep MTF LQRWQGLADNKICAFWFVIRRFSEERV PQAAASMTFTLLALVPVLTVMVAVASIF
   orf144-1      MTF LQRWQGLADNKICAFWFVIRRFSEERV PQAAASMTFTLLALVPVLTVMVAVASIF

10 orf144ng-1.pep PVFDRWSDSFVSFVNQTI V PQADMVFDYIDAFRDQANRLTAIGSVMLVVTSLMLIRTID
   orf144-1      PVFDRWSDSFVSFVNQTI V PQADMVFDYIDAFRDQANRLTAIGSVMLVVTSLMLIRTID

15 orf144ng-1.pep NAFNRIWRVNTQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVD SVLSSGAQQWADAL
   orf144-1      NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVD AALASGAPQWSGAL

20 orf144ng-1.pep KTAARLAFMTLLWGLYRFV PNRFPARQAFV GALITAFCL ETARFLFTWYMG NFDGYRS
   orf144-1      RTAATLTFTLLWGLYRFV PNRFPARQAFV GALITAFCL ETARSLFTWYMG NFDGYRS

25 orf144ng-1.pep IYGAF AAVPFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILLLL
   orf144-1      IYGAF AAVPFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILLLL

   orf144ng-1.pep DAAQKEGRTLSVQEFRRHINMGYDELGE LLEKLARYGYIYSGRQGWVLTGADSIELSEL
   orf144-1      DAAQKEGKALPVQEFRRHINMGYDELGE LLEKLARHGYYIYSGRQGWVLTGADSIELNEL

   orf144ng-1.pep FKL FVYRPLPVERDHVNQAVDAVMTPCLQTLNM TLAEFDAQAKKQQQS
   orf144-1      FKL FVYRPLPVERDHVNQAVDAVMTPCLQTLNM TLAEFDAQAKKRQ

```

On this basis of this analysis, including the identification of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### 35 Example 75

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 629>:

```

1  ..AGACACGCC GCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA
51 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
101 GCACCGATAT GCGTCAGGAA ATTTCCGCC TCCTCATCCT GCTGCAACGC
40 151 ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCGCCAAAG
   201 CTGCTTGAA ACACGGGAAC ACGGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 630; ORF146>:

```

1  ..RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTDMRQE ISALVILLQR
51 TRRKWLDAHE RQHLRQSLLE TREHG*

```

45 Further work revealed the complete nucleotide sequence <SEQ ID 631>:

```

1  ATGAACACCT CGCAACGCAA CCGCCTCGTC AGCCGCTGGC TCAACTCCTA
51 CGAACGCTAC CGCTACCGCC GCCTCATCCA CGCCGTCGGG CTCGGCGGGG
101 CCGTCTGTGT CGCCACCGCC TCCGCCGGC TGCTCCACCT CCAACACGGC
50 151 GAGTGGATAG GGATGACCGT CTTCTGTCGTC CTCGGCATGC TCCAGTTTCA
   201 AGGGGCGATT TACTCCAAGG CGGTGGAACG TATGCTCGGC ACGGTCATCG
   251 GGCTGGGCGC GGGTTTGGGC GTTTTATGGC TGAACCAGCA TTATTCCAC
   301 GGCAACCTCC TCTTCTACCT CACCGTCGGC ACGGCAAGCG CACTGGCCGG
   351 CTGGGCGGCG GTCGGCAAAA ACGGCTACGT CCCTATGCTG GCAGGGCTGA
   401 CGATGTGTAT GCTCATCGGC GACAACGGCA GCGAATGGCT CGACAGCGGA
55 451 CTCATGCGCG CCATGAACGT CCTCATCGGC GCGGCCATCG CCATCGCCGC

```

-352-

5 501 CGCCAAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG  
 551 CCGACAACCT GGCCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC  
 601 AGGCGCATGA CCCGCGAAGC CCTCGAGGAG AACATGGCGA AAATGCGCCA  
 651 AATCAACGCA CGCATGGTCA AAAGCCGAG CCATCTCGCC GCCACATCGG  
 701 GCGAAAGCCG CATCAGCCCC GCCATGATGG AAGCCATGCA GCACGCCAC  
 751 CGTAAATCG TCAACACCAC CGAGCTGCTC CTGACCACCG CCGCCAAGCT  
 801 GCAATCTCCC AACTCAACG GCAGCGAAAT CCGGCTGCTT GACCGCCACT  
 851 TCACACTGCT CCAAACCGAC CTGCAACAAA CCGTCGCCCT TATCAACGGC  
 10 901 AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA  
 951 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA  
 1001 GCACCAATAT GCGTCAGGAA ATTTCCGCCC TCGTCATCCT GCTGCAACGC  
 1051 ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCGCCAAAG  
 1101 CCTGCTTGAA ACACGGGAAC ACGGCTGA

This corresponds to the amino acid sequence <SEQ ID 632; ORF146-1>:

15 1 MNTSQNRNLV SRWLNSYERY RYRRLIHAVR LGGAVLFATA SARLLHLQHG  
 51 EWIGMTVFVV LGMLQFQGA YSKAVERMLG TVIGLGAGLG VLWLNQHYFH  
 101 GNLLFYLTVG TASALAGWAA VGKNGYVPEML AGLTMCMLIG DNGSEWLDG  
 151 LMRAMNVLIG AAIATAAAL LPLKSTLMWR FMLADNLADC SKMIAEISNG  
 20 201 RRMTRERLEE NMAKMRQINA RMVKSRSILA ATSGESRISP AMMEAMQHAH  
 251 RKIVNTTELL LTTAAKLQSP KLNGSEIRLL DRHFTLLQTD LQQTVALING  
 301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMROE ISALVILLQR  
 351 TRRWLDAHE RQHLRQSLLE TREHG\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF146 shows 98.6% identity over a 74aa overlap with an ORF (ORF146a) from strain A of *N. meningitidis*:

	orf146.pep			10	20	30
				RHARRIRIDTAINPELEALAEHLHYQWQGF		
30	orf146a	KLNGSEIRLLDRHFTLLQTDLQQTVALINGRHARRIRIDTAINPELEALAEHLHYQWQGF	280	290	300	310
						320
						330
	orf146.pep	LWLSTDMRQEISALVILLQRTTRRWLDAHERQHLRQSLLETREHG	40	50	60	70
35	orf146a	LWLSTNMROEISALVILLQRTTRRWLDAHERQHLRQSLLETREHSX	340	350	360	370

The complete length ORF146a nucleotide sequence <SEQ ID 633> is:

40 1 ATGAACACCT CGCAACGCAA CGCCTCGTC AGCCGCTGGC TCAACTCCTA  
 51 CGAACGCTAC CGCTACCGCC GCCTCATCCA CGCCGTCCGG CTCGGCGGGG  
 101 CCGTCTGTG CGCCACCGCC TCCGCCCGGC TGCTCCACCT CCAACACGGC  
 151 GAGTGGATAG GGATGACCGT CTTCTGTCGTC CTCGGCATGC TCCAGTTTCA  
 201 AGGGGCGATT TACTCCAAGG CGGTGGAACG TATGCTCGGC ACGGTCATCG  
 45 251 GGCTGGGCGC GGGTTTGGGC GTTTTATGGC TGAACCAGCA TTATTTCAC  
 301 GGCAACCTCC TCTTCTACCT CACCGTCGGC ACGGCAAGCG CACTGGCCGG  
 351 CTGGGCGGCG GTCGGCAAAA ACGGCTACGT CCCTATGCTG GCGGGGCTGA  
 401 CGATGTGCAT GCTCATCGGC GACAACGGCA GCGAATGGT CGACAGCGGC  
 451 CTGATGCGCG CGATGAACGT CCTCATCGGC GCGGCCATCG CCATCGCCGC  
 50 501 CGCCAAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG  
 551 CCGACAACCT GACCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC  
 601 AGGCGCATGA CCCGCGAAGC CCTCGAAGAG AACATGGCGA AAATGCGCCA  
 651 AATCAACGCA CGCATGGTCA AAAGCCGAG CCACCTCGCC GCCACATCGG  
 701 GCGAAAGCCG CATCAGCCCC GCCATGATGG AAGCCATGCA GCACGCCAC  
 751 CGTAAATCG TCAACACCAC CGAGCTGCTC CTGACCACCG CCGCCAAGCT  
 55 801 GCAATCTCCC AACTCAACG GCAGCGAAAT CCGGCTGCTT GACCGCCACT  
 851 TCACACTGCT CCAAACCGAC CTGCAACAAA CCGTCGCCCT TATCAACGGC  
 901 AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA  
 951 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA  
 1001 GCACCAATAT GCGTCAGGAA ATTTCCGCCC TCGTCATCCT GCTGCAACGC  
 60 1051 ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCGCCAAAG  
 1101 CCTGCTTGAA ACACGGGAAC ACAGTTGA

This encodes a protein having amino acid sequence <SEQ ID 634>:

```

1  MNTSQNRNLV SRWLNSYERY RYRRLIHAVR LGGAVLFATA SARLLHLQH
51 EWIGMTVFV LGMLQFQGA YSKAVERMLG TVIGLGAGLG VLWLNQHYFH
101 GNLLFYLTVG TASALAGWAA VGKNGYVPM LAGLTMCLIG DNGSEWFD
5  151 LMRAMNVLIG AAIAIAAKL LPLKSTLMWR FMLADNLTD SKMIAEISNG
201 RRMTRERLEE NMAKMRQINA RMVKSRSHLA ATSGESRISP AMMEAMQHAH
251 RKIVNTTELL LTAAKLQSP KLGSEIRLL DRHFTLLQTD LQQTVALING
301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR
351 TRRKWLDAHE RQHLRQSLE TREHS*

```

10 ORF146a and ORF146-1 show 99.5% identity in 374 aa overlap:

```

orfl46a.pep  MNTSQNRNLVSRWLNSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFV
orfl46-1      MNTSQNRNLVSRWLNSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFV
15 orfl46a.pep  LGMLQFQGAISKKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTVG
orfl46-1      LGMLQFQGAISKKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTVG
20 orfl46a.pep  VGKNGYVPM LAGLTMCLIGDNGSEWFD SGLMRAMNVLIGAAIAIAAKLLPLKSTLMWR
orfl46-1      VGKNGYVPM LAGLTMCLIGDNGSEWLD SGLMRAMNVLIGAAIAIAAKLLPLKSTLMWR
25 orfl46a.pep  FMLADNLTDCKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP
orfl46-1      FMLADNLADCKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP
30 orfl46a.pep  AMMEAMQHAHRKIVNTTELLLTAAKLQSPKLGSEIRLLDRHFTLLQTDLQQTVALING
orfl46-1      AMMEAMQHAHRKIVNTTELLLTAAKLQSPKLGSEIRLLDRHFTLLQTDLQQTVALING
35 orfl46a.pep  RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQRTTRRKWLDAHE
orfl46-1      RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQRTTRRKWLDAHE
orfl46a.pep  RQHLRQSLE TREHSX
orfl46-1      RQHLRQSLE TREHGX

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

40 ORF146 shows 97.3% identity over a 75aa overlap with a predicted ORF (ORF146ng) from *N.gonorrhoeae*:

```

orfl46.pep  RHARRIRIDTAINPELEALAEHLHYQWQGF 30
orfl46ng    KLGSEIRLLDRHFTLLQTDLQQTAA LINGRHARRIRIDTAINPELEALAEHLHYQWQGF 364
45 orfl46.pep  LWLSTDMRQEISALVILLQRTTRRKWLDAHERQHLRQSLE TREHG 75
orfl46ng    LWLSTNMRQEISALVIPLQRTTRRKWLDAHERQHLRQSLE TREHG 409

```

An ORF146ng nucleotide sequence <SEQ ID 635> was predicted to encode a protein having amino acid sequence <SEQ ID 636>:

```

1  MSGVRFPSA PIPSTDPPSG SLCCFTFPLQ TASDMNSSQR KRLSGRWLNS
51 YERYRHRRLI HAVRLGGTVL FATALARLLH LQHGEWIGMT VFVVLGMLQF
101 QGAIYSNAVE RMLGTVIGL AGLGVLWLNQ HYFHGNLLFY LTIGTASALA
55 151 GWAAVGKNGY VFMLAGLTM MLIGDNGSEW LDSGLMRAMN VLIGAAIAIA
201 AAKLLPLKST LMWRFMLADN LADCSKMAIE ISNGRRMTRE RLEQNMVKMR
251 QINARMVKSR SHLAATSGES RISPSMMEAM QHAHRKIVNT TELLTTAAK
301 LQSPKLGSE IRLDRHFTL LQTDLQQTAA LINGRHARRI RIDTAINPEL
351 EALAEHLHYQ WQGF LWLSTN MRQEISALVI PLQRTTRKWL DAHERQHLRQ
401 SLLE TREHG*

```

60 Further work revealed the following gonococcal DNA sequence <SEQ ID 637>:

1 ATGAAGCTCT CGCAACGCAA ACGCCTTTCC GgccGCTGGC TCAACTCCTA  
 51 CGAAGCTac cGCCaccGCC GCCTCATACA TGCCGTGCGG CTCGGCGgaa  
 101 ccgtCCTGTT CGCCACCGCA CTCGCCGgc tACTCCACCT CCAacacggc  
 151 gAATGGATAG GGatgaCCGT CTTCGTGCTC CTCGGCATGC TCCAGTTCCA  
 201 AGGCgcgatt tActccaacg cgggtGAacg taTGctcggg acggtcatcg  
 251 ggctgGGCGC GGGTTTGGgc gTTTTATGGC TGAACCAGCA TTAttccac  
 301 ggcaacCTcc tcttctacct gaccatcggc acggcaagcg cactggccgg  
 351 ctGGGCGGCG GTCCGCAAAA acggctacgt ccctatgctg GCGGGGctgA  
 401 CGATGTGCAT gctcatcggc gACAACGGCA GCGAATGGCT CGACAGCGGC  
 451 CTGATGCGCG CGATGAACGT CCTCATCGGC GCCGCCATCG CCATTGCCGC  
 501 CGCCAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG  
 551 CCGACAACCT GGCCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC  
 601 AGGCGTATGA CGCGCGAACG TTTGGAGCAG AATATGGTCA AAATGCGCCA  
 651 AATCAACGCA CGATGGTCA AAAGCCGCGC CCACCTCGCC GCCACATCGG  
 701 GCGAAAGCCG CATCAGCCCC TCCATGATGG AAGCCATGCA GCACGCCAC  
 751 CGCAAAATCG TCAACACCAC CGAGCTGCTC CTGACCACCG CCGCCAAGCT  
 801 GCAATCTCCC AAACCTCAAC GCAGCGAAAT CCGGCTGCTC GACCGCCACT  
 851 TCACACTGCT CCAAACCGAC CTGCAACAAA CCGCCGCCCT CATCAACGGC  
 901 AGACACGCCG GCCGCATCCG CATCGACACC GCCATCAACC CCGAAGTGA  
 951 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA  
 1001 GCACCAATAT GCGTCAGGAA ATTTCCGCC TCGTCATCCT GCTGCAACGC  
 1051 ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TCGCCCAAG  
 1101 CCTGCTTGA ACACGGGAAC ACGGCTGA

This corresponds to the amino acid sequence <SEQ ID 638; ORF146ng-1>:

1 MNSSQRKRLS GRWLSYERY RHRRLIHAVR LGGTVLFATA LARLLHLQHG  
 51 EWIGMTVFVV LGMLOFQGA YSNAVERMLG TVIGLGAGLG VLWLNQHYFH  
 101 GNLLFYLTIG TASALAGWAA VGKNGYVPEML AGLTMCMLIG DNGSEWLD SG  
 151 LMRAMNVLIG AATAIAAAKL LPLKSTLMWR FMLADNLAD SKMIAEISNG  
 201 RRMTRERLEQ NMVKMRQINA RMVKSRSHLA ATSGESRISP SMMEAMQHAH  
 251 RKIVNTELL LTTAAKLQSP KLNGSEIRLL DRHFTLLQTD LQQTAAALING  
 301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR  
 351 TRRWLDAHE RQHLRQSLE TREHG\*

ORF146ng-1 and ORF146-1 show 96.5% identity in 375 aa overlap

35 orf146-1.pep MNTSQRNRLVSRWLSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHG  
 orf146ng-1 MNSSQRKRLSGRWLSYERYRHRRLIHAVRLGGTVLFATALARLLHLQHG  
 40 orf146-1.pep LGMLOFQGAISKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAA  
 orf146ng-1 LGMLOFQGAISNAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAA  
 45 orf146-1.pep VGKNGYVPEMLAGLTMCMLIGDNGSEWLD SGLMRAMNVLIGAAIAAAKLLPLKSTLMWR  
 orf146ng-1 VGKNGYVPEMLAGLTMCMLIGDNGSEWLD SGLMRAMNVLIGAAIAAAKLLPLKSTLMWR  
 50 orf146-1.pep FMLADNLADCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP  
 orf146ng-1 FMLADNLADCSKMIAEISNGRRMTRERLEQNMVKMRQINARMVKSRSHLAATSGESRISP  
 55 orf146-1.pep AMMEAMQHAHRKIVNTELLLTAAKLQSPKLNGSEIRLLDRHFTLLQTDLQQTVALING  
 orf146ng-1 SMMEAMQHAHRKIVNTELLLTAAKLQSPKLNGSEIRLLDRHFTLLQTDLQQTAAALING  
 60 orf146-1.pep RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQTRRWLDAHE  
 orf146ng-1 RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQTRRWLDAHE  
 orf146-1.pep RQHLRQSLE TREHG  
 orf146ng-1 RQHLRQSLE TREHG

Furthermore, ORF146ng-1 shows homology with a hypothetical *E.coli* protein:

sp|P33011|YEEA ECOLI HYPOTHETICAL 40.0 KD PROTEIN IN COBU-SBMC INTERGENIC REGION  
 >gi|1736674|gnl|PID|d1016553 (D90838) ORF ID:o348#20; similar to [SwissProt  
 Accession Number P33011] [Escherichia coli] >gi|1736682|gnl|PID|d1016560 (D90839)  
 ORF ID:o348#20; similar to [SwissProt Accession Number P33011] [Escherichia coli]

-355-

>gi|1788318 (AE000292) f352; 100% identical to fragment YEEA\_ECOLI SW: P33011 but has 203 additional C-terminal residues [Escherichia coli] Length = 352  
Score = 109 bits (271), Expect = 2e-23  
Identities = 89/347 (25%), Positives = 150/347 (42%), Gaps = 21/347 (6%)

Query: 20 YRHRRLIHAVRLGGTVLFFATALARLLHLOHGEWIGMTVFVVLGMLQFQGAISNAVERML 79  
YRH R++H R+ L + RL + W +T+ V++G + F G + A ER+  
Sbjct: 15 YRHYRIVHGTRVALAFLTLFLIIRLFTIPESTWPLVTMVIMGPISFWGNVVPRAFERIG 74

Query: 80 GTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAAVGKNGYVPMLAGLTMCLMI 139  
GTV+G GL L L L + A L GW A+GK Y +L G+T+ +++  
Sbjct: 75 GTVLGSILGLIALQLE---LISLPLMLVWCAAAMFLCGWLALGKKPYQGLLIGVTLAIVV 131

Query: 140 GDNGSEWLDSGLMRAMNVLIGXXXXXXXXXXKLLPLKSTLMWREMLADNLADCSKMIAEISN 199  
G E +D+ L R+ +V++G + P ++ + WR LA +L + +++ +  
Sbjct: 132 GSPTGE-IDTALWRSRGDVILGSLLAMLETGIWPQRAFIHWRIQLAKSLTEYNRVYQSASF 190

Query: 200 GRRMTRERLEQNMVKMRQINARMVKSRSLSHAATSGESRISPSMMEAMQHAHRKIVNXXXX 259  
+ R RLE ++ K+ VK R +A S E+RI S+ E +Q +R +V  
Sbjct: 191 PNLLERPRLESHLQKLL---TDAVKMRGLIAPASKETRIKSIYEGIQITINRNLCMLEL 247

Query: 260 XXXXXXXXQSPK---LNGSEIRLLDRHFXXXXXXXXXXAALINGRHARRIRIDTAINPEL 316  
+ LN ++R D AL G +N +  
Sbjct: 248 QINAYWATRP SHFVLLNAQKLR--DTQHM MQILLSLVHALYEGNPQPVFANTEKLNDV 305

Query: 317 EALAEHL--HYQWQ-----GFLWLSTNMQRQEISALVILLQRTTRK 354  
E L + L H+ + G++WL+ ++ L L+ R RK  
Sbjct: 306 EELRQLLNHHDLKVVETPIYGYVWLNMETAHQLELLSNLICRALRK 352

On the basis of this analysis, including the identification of several transmembrane domains in the  
gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and  
their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 76

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 639>

1 ..GCCGAAGACA CGCGCGTTAC CGCACAGCTT TTGAGCGCGT ACGGCATTCA  
51 GGGCAAATC GTCAGTGTGC GCGAACACAA CGAACGGCAG ATGGCGGACA  
101 AGATTGTTCG CTATCTTTCA GACGGCATGG TTGTGGCACA GGTTCCGAT  
151 GCGGGTACGC CGGCCGTGTG CGACCCGGGC GCGAACTCG CCCGCCGCGT  
201 GCGTGAGGCC GGGTTTAAAG TCGTTCCCGT CGTGGGCGCA AC.GCGGTGA  
251 TGGCGGCTTT GAGCGTGGCC GGTGTGGAAG GATCCGATTT TTATTTCAAC  
301 GGTTTTGTAC CGCCGAAATC GGGAGAACGC AGGAACTGT TTGCCAAATG  
351 GGTGCGGGCG GCGTTTCTTA TCGTCATGTT TGAAACGCCG CACCGCATCG  
401 GTGCAGCGCT TGCCGATATG GCGGAAGTGT TCCCCGAACG CCGATTAATG  
451 CTGGCGCGCG AAATTACGAA AACGTTTGAA ACGTTCTTAA GCGGCACGGT  
501 TGGGGAAT CAGACGGCAT TGTCTGCCGA CGGCGACCAA TCGCGCGGCG  
45 551 AGATGGTGT GGTGCTTTAT CCGGCGCAGG ATGAAAAACA CGAAGGCTTG  
601 TCCGAGTCCG CGCAAAACAT CATGAAATC CTCACAGCCG AGCTGCCGAC  
651 CAAACAGGCG GCGGAGCTTG CTGCCAAAAT CACGGGCGAG GGAAAGAAAG  
701 CTTTGTACGA T..

This corresponds to the amino acid sequence <SEQ ID 640; ORF147>:

1 ..AEDTRVTAQL LSAYGIQGLK VSVREHNERQ MADKIVGYLS DGMVVAQVSD  
51 AGTFAVCDPG AKLARRVREA GFKVVPVVG XAVMAALSVA GVEGSDFYFN  
101 GFVFPKSGER RKLFAKWVRA AFPIVMFETP HRIGALADM AELFPERRLM  
151 LAREITKTFE TFLSGTVGEI QTALSADGDQ SRGEMVLVLY PAQDEKHEGL  
201 SESAQNIMKI LTAELPTKQA AELAAKITGE GKKALYD..

Further work revealed the complete nucleotide sequence <SEQ ID 641>:

1 ATGTTTCAGA AACATTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC  
51 ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCGGAC ATTACCTGC  
101 GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG  
151 CGCGTTACCG CACAGCTTTT GAGCGCGTAC GGCATTACAG GCAAACCTCGT



-356-

201 CAGTGTGCGC GAACACAACG AACGGCAGAT GGCGGACAAG ATTGTGCGCT  
 251 ATCTTTTCAGA CGGCATGGTT GTGGCACAGG TTTCCGATGC GGGTACGCCG  
 301 GCCGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GTGAGGCCGG  
 351 GTTTAAAGTC GTTCCCGTCG TGGGCGCAAG CGCGGTGATG GCGGCTTTGA  
 401 GCGTGGCCGG TGTGGAAGGA TCCGATTTT ATTCAACGG TTTTGTACCG  
 451 CCGAAATCGG GAGAACGCAG GAAACTGTTT GCCAAATGGG TCGGGCGGGC  
 501 GTTTCCTATC GTCATGTTTG AAACGCCGCA CCGCATCGGT GCGACGCTTG  
 551 CCGATATGGC GGAACGTGTC CCCGAACGCC GATTAATGCT GGCGCGCGAA  
 601 ATTACGAAAA CGTTGAAAC GTTCTTAAGC GGCACGGTTG GGGAAATTCA  
 651 ACGGCATTG TCTGCCGACG GCAACCAATC GCGCGCGAG ATGGTGTGG  
 701 TGCTTTATCC GCGCAGGAT GAAAAACAG AAGGCTTGTC CGAGTCCGCG  
 751 CAAAACATCA TGAATACTCT CACAGCCGAG CTGCCGACCA AACAGGCCGC  
 801 GGAGCTTGCT GCCAAAATCA CGGGCGAGGG AAAGAAAGCT TTGTACGATC  
 851 TGGCTCTGTC TTGAAAAAC AAATAG

15 This corresponds to the amino acid sequence <SEQ ID 642; ORF147-1>:

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT  
 51 RVTAQLLSAY GIQKLVSVR EHNERQMA DK IVGYLSDGMV VAQVSDAGT P  
 101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALS VAGVEG SDFYFNGFVP  
 151 PKSGERRKLF AKWVRAAFPI VMFETPHRIG ATLDMAELF PERRLMLARE  
 201 ITKTFETFLS GTVGEIQTAL SADGNQSRGE MVLVLYPAQD EKHEGLSESA  
 251 QNIMKILTAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K\*

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical protein ORF286 of *E. coli* (accession number U18997)

ORF147 and *E. coli* ORF286 protein show 36% aa identity in 237aa overlap:

25 Orf147: 1 AEDTRVTAQLLSAYGIQKLVSVREHNERQMA DKIVGYLSDGMVVAQVSDAGT PAVCDPG 60  
 AEDTR T LL +GI +L ++ +HNE+Q A+ ++ L +G +A VSDAGT P + DPG  
 Orf286: 43 AEDTRHTG LLLQHFGINARLFALHDHNEQQKAETLLAKLQEGQNIALVSDAGT PLINDPG 102  
 30 Orf147: 61 AKLARRVREXXXXXXXXXXXXXXXXXXXXXXXXXEGSDFYFNGFVPPKSGERRKLF AKWVRA 120  
 L R RE F + GF+P KS RR  
 Orf286: 103 YHLVRTCREAGIRVVP LPGPCAAITALSAGLP SDRFCYEGFLPAKSKGRRDALKAIEAE 162  
 Orf147: 121 AFPIVMFETPHRIGAAALADMAELFPERR-LMLAREITKTFETFLSGTVGEIQTALSADGD 179  
 ++ +E+ HR+ +L D+ + E R ++LARE+TKT+ET VGE+ + D +  
 35 Orf286: 163 PRTLIFYESTHRLDSLEDIVAVLGESRYVVLARELT KTWTETHGAPVGE LLA WVKEDEN 222  
 Orf147: 180 QSRGEMVLVLYPAQDEKHEGLSESAQNIMKILTAE LPTKQAAELA AKITGEGKKALY 236  
 + +GEMVL++ + E L A + +L AELP K+AA LAA+I G K ALY  
 40 Orf286: 223 RRGEMVLIV-EGHKAQEDLPADALRT LALLQAE LPLKKAALAAEIHGVKKNALY 278

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF147 shows 96.6% identity over a 237aa overlap with ORF75a from strain A of *N. meningitidis*:

45 orf147.pep AEDTRVTAQLLSAYGIQKLVSVREHNERQ  
 orf75a TLYVVATPIGNLADITLRLALAVLQKADIICAEDTRVTAQLLSAYGIQKLVSVREHNERQ  
 20 30 40 50 60 70  
 50 orf147.pep MADKIVGYLSDGMVVAQVSDAGT PAVCDPGAKLARRVREAGFKVVPVVGAXAVMAALSVA  
 orf75a MADKIVGYLSDGMVVAQVSDAGT PAVCDPGAKLARRVREAGFKVVPVVGASAVMAALSVA  
 80 90 100 110 120 130  
 55 orf147.pep GVEGSDFYFNGFVPPKSGERRKLF AKWVRAAFPIVMFETPHRIGAAALADMAELFPERRLM  
 orf75a GVAGSDFYFNGFVPPKSGERRKLF AKWVRAAFPIVMFETPHRIGATLADMAELFPERRLM  
 140 150 160 170 180 190  
 60 orf147.pep LAREITKTFETFLSGTVGEIQTALSADGDQSRGEMVLVLYPAQDEKHEGLSESAQNIMKI

orf75a  
LAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLVPAQDEKHEGLSESAQNIMKI  
200 210 220 230 240 250

5

	220	230
orf147.pep	LTAELPTKQAAELAAKITGEGKKALYD	
orf75a	LTAELPTKQAAELAAKITGEGKKALYDLALSWKXKX	
	260	270 280 290

### Homology with a predicted ORF from *N.gonorrhoeae*

15	orf147.pep	AEDTRVTAQLLSAYGIQGLVSVREHNERQ	30
	orf147ng	TLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQGLVSVREHNERQ	85
20	orf147.pep	MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGAXAVMAALSVA	90
	orf147ng	MADKIVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGASAVMAALSVA	145
25	orf147.pep	GVEGSDFYFNGFVPPKSGERRKLEAKWVRAAFPVIMFETPHRIGAAALDMAELFPERRLM	150
	orf147ng	GVAESDFYFNGFVPPKSGERRKLEAKWVRAAFPVVMFETPHRIGATLADMAELFPERRLM	205
30	orf147.pep	LAREITKTFTFSLSGTVGEIQTALSADGDQSRGEMVLVLVPAQDEKHEGLSESAQNIMKI	210
	orf147ng	LAREITKTFTFSLSGTVGEIQTALAADGNQSRGEMVLVLVPAQDEKHEGLSESAQNAMKI	265
30	orf147.pep	LTAELPTKQAAELAAKITGEGKKALYD	237
	orf147ng	LAELPTKQAAELAAKITGEGKKALYDLALSWKNK	300

35	1	MSVFQTAFFM	FQKHLQKASD	SVVGGTLYV	ATPIGNLADI	TLRALAVLQK
	51	ADIICAEDTR	VTAQLLSAYG	IQGRLVSVRE	HNERQMAKV	IGFLSDGLVV
	101	AQVSDAGTPA	VCDPGAKLAR	RVREAGFKV	PVVGASAVMA	ALSVAGVAES
	151	DFYFNGFVPP	KSGERRKLFA	KWVRAAFPV	MFETPHRIGA	TLADMAELFP
	201	ERRIMLAREI	TKTFETFLSG	TVGEIQTALA	ADGNQSRGEM	VLVLYPAQDE
40	251	KHEGLSESAQ	NAMKILAAEL	PTQAAELAA	KITGEGKKAL	YDLALSWKNK
	301	*				

	1	ATGTTTCAGA	AACACTTGCA	GAAAGCCTCC	GACAGCGTCG	TCGGAGGGAC
45	51	ATTATACGTG	GTTGCCACGC	CCATCGGCAA	TTTGGCAGAC	ATTACCCTGC
	101	GCGCTTTGGC	GGTATTGCAA	AAGGCGGACA	TCATTGTGCG	CGAAGACACG
	151	CGCGTTACTG	CGCAGCTTTT	GAGCGCGTAC	GGCATTGAGC	CGAGGTGTGT
	201	CAGTGTGCGC	GAACACAACG	AGCGGCAGAT	GGCGGACAAG	GTAATCGGTT
	251	TCCTTTCAGA	CGGCCTGGTT	GTGGCGCAGG	TTTCCGATGC	GGGTACGCCG
50	301	GCCGTGTGCG	ACCCGGGGCG	GAAACTCGCC	CGCGCGGTGC	GCGAAGCAGG
	351	GTTCAAAGTC	GTTCCCGTCG	TGGGCGCAAG	CGCGGTAATG	GCGGCGTGTG
	401	GTTGTGGCCGG	TGTGGCGGAA	TCCGATTTT	ATTTCAACGG	TTTGTGTACG
	451	CCGAAATCGG	GCGAACGTAG	GAAATTGTTT	GCCAAATGGG	TGCGGGCGGC
	501	ATTTCTGTGC	GTCATGTTTG	AAACGCCGCA	CCGAATCGGG	GCAACGCTTG
55	551	CCGATATGGC	GGAATTGTTT	CCCGAACCGC	GTCATGATCT	GGCGCGCGAA
	601	ATACAGAAAA	CGTTTGAAC	GTCTTAAAG	GCGACGGTTG	GGGAAATTCA
	651	GACGGCATTG	GCGGCGGACG	GCAACCAATC	GCGCGGCGAG	ATGGTGTGTG
	701	TGCTTTATCC	GGCGCAGGAT	GAAAACACG	AAGGCTTGTC	CGAGTCTGCG
	751	CAAAATGCGA	TGAAAACTCT	TGCGGCCGAG	CTGCCGACCA	AGCAGGCGGC
60	801	GGAAGTTGCG	GCCAGATTA	CAGGTGAGGG	CAAAAAGGCT	TTGTACGATT
	851	TGGCACTGTC	GTGGAATAAC	AAATGA		

This corresponds to the amino acid sequence <SEQ ID 646; ORF147ng-1>:

```

      1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
      51 RVT AQLLSAY GIQGRIVSVR EHNERQMADK VIGFLSDGLV VAQVSDAGTP
      101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVAE SDFYFNGFVP
      151 PKSGERRKLF AKWVRAAFPV VMFETPHRIG ATLADMAELF PERRIMLARE
      201 ITKTFETFLS GTVGEIQTAL AADGNQSRGE MVLVLYPAQD EKHEGLSESA
      251 QNAMKILAAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K*
  
```

ORF147ng shows homology to a hypothetical *E.coli* protein:

```

10  sp|P45528|YRAL_ECOLI HYPOTHETICAL 31.3 KD PROTEIN IN AGAI-MTR INTERGENIC REGION
    (F286)
    >gi|606086 (U18997) ORF_f286 [Escherichia coli]
    >gi|1789535 (AE000395) hypothetical 31.3 kD protein in agai-mtr intergenic region
    [Escherichia coli] Length = 286
    Score = 218 bits (550), Expect = 3e-56
15  Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)

    Query: 4   KHLQKASDSVVGGTLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQ 63
              K Q A +S G LY+V TPIGNLADIT RAL VLQ D+I AEDTR T LL +GI
    Sbjct: 2   KQHQSADNSQ--GQLYIVPTPIGNLADITQRALEVLQAVDLIAAEDTRHTGLLLQHFGIN 59

    Query: 64   GRLVSVREHNERQMADKVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPV 123
              RL ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG L R REAG +VVP+
    Sbjct: 60   ARLFALHDHNEQQKAETLLAKLQEGQNIALVSDAGTFLINDPGYHLVRTCREAGIRVVP 119

    Query: 124  VGASAVMAALS VAGVAESDFYFNGFVPPKSGERRKLF AKWVRAAFPVVMFETPHRIGATL 183
              G A + ALS AG+ F + GF+P KS RR ++ +E+ HR+ +L
    Sbjct: 120  PGPCAAITALS AAGLPSDRFCYEGFLPAKSKGRRDALKAI EAEPRTLIFYESTHRLDLSL 179

    Query: 184  ADMAELFPERR-LMLAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEK 242
              D+ + E R ++LARE+TKT+ET VGE+ + D N+ +GEMVL++ +
    Sbjct: 180  EDIVAVLGESRYVVLARELTKTWETIHGAPVGE LLAWVKEDENRRKGMVLIIV-EGHKAQ 238

    Query: 243  HEGLSESAQNAMKILAAELPTKQAAELA AKITGEGKKALYDLAL 286
              E L A + +L AELP K+AA LAA+I G K ALY AL
    Sbjct: 239  EEDLPADALRTLALLQAEPLKKAALAAEIHGVKNALYKYAL 282
  
```

Based on the computer analysis and the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 77

40 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 647>

```

      1 ATGAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCGAA
      51 AACCGGTGCG ATCCGCTTCT C.GCTGCTTA CTTAGCCATA TGCCTGTCGT
      101 TCGGCATTCT TCCCAAGCC TGGGCGGGAC ACACTTATTT CGGCATCAAC
      151 TACCAATACT ATCGCGACTT TGCCGAAAAT AAAGGCAAGT TTGCAGTCGG
      201 GCGGAAAGAT ATTGAGGTTT ACAACAAAAA AGGGGAGTTG GTCGGCAAAT
      251 CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTG GCGTAACGGC
      301 GTGGCGGCAT TGGTGGGCGT ATCAATATAT TGTGAGCGTG GCACATAACG
      351 GCGGCTATAA CAACGTTGAT TTTGGTGC GG AAGGAak.AA tATCCC.GAT
      401 CAACAACGw TTACTTATAA AATTGTGAAA CGGAATAATT ATAAAGCAGG
      451 GACTAAAGGC CATCCTTATG GCGGCGATTA TCATATGCCG CGTTTGCGTA
      501 AATWTGTCAC AGATGCAGAA CCTGTTGAAA TGACCAGTTA TATGGATGGG
      551 CGGAAATATA TCGATCAAAA TAATTACCCT GACCGTGTTC GTATTGGGGC
      601 AGGCAGGCAA TATTGGCGAT CTGATGAAGA TGAGCCCAAT AACCGCGAAA
      651 GTTCATATCA TATTGCAAGT .....
      701 ..... GGCTC ACCAATGTTT ATCTATGATG CCCAAAAGCA
      751 AAAGTGGTTA ATTAATGGGG TATTGCAAAAC GGGCAACCCC TATATAGGAA
      801 AAAGCAATGG CTTCAGCTG GTTCGTAAAG ATTGGTTCTA TGATGAAATC
      851 TTTGCTGGAG ATACCCATTC AGTATTCTAC GAACCACGTC AAAATGGGAA
      901 ATACTCTTTT AACGACGATA ATAATGGCAC AGGAAAAATC AATGCCAAAC
  
```

-359-

951 ATGAACACAA TTCTCTGCCT AATAGATTAA AAACACGAAC CGTTCAATTG  
 1001 TTTAATGTTT CTTTATCCGA GACAGCAAGA GAACCTGTTT ATCATGCTGC  
 1051 AGGTGGTGTC AACAGTTATC GACCCAGACT GAATAATGGA GAAAATATTT  
 1101 CCTTTATTGA CGAAGGAAAA GGCGAATTGA TACTTACCAG CAACATCAAT  
 5 1151 CAAGGTGCTG GAGGATTATA TTTCCAAGGA GATTTTACGG TCTCGCCTGA  
 1201 AAATAACGAA ACTTGGCAAG GCGCGGGCGT TCATATCAGT GAAGACAGTA  
 1251 CCGTTACTTG GAAAGTAAAC GGCGTGGCAA ACGACCGCCT GTCCAAAATC  
 1301 GGCAAAGGCA CGCTG.....  
 //  
 10 2101 ..... GATAAAG  
 2151 TGACTGCTTC ATTGACTAAG ACCGACATCA GCGGCAATGT CGATCTTGCC  
 2201 GATCAGCTC ATTTAAATCT CACAGGGCTT GCCACACTCA ACGGCAATCT  
 2251 TAGTGCAAAT GGCGATACAC GTTATACAGT CAGCCACAAC GCCACCCAAA  
 15 2301 ACGGCAACCK TA<sub>9</sub>CCtCGtG G.sAATGcCC AAGCAACATT TAATCAAGCC  
 2351 ACATTAAACG GCAACACATC GGCTTCgGGC AATGCTTCAT TTAATCTAAG  
 2401 CGACCACGCC GTACAAAACG GCAGTCTGAC GCTTTCCGGC AACGCTAAGG  
 2451 CAAACGTAAAG CCATTCCGCA CTCACGGTA ATGTCTCCCT AGCCGATAAG  
 2501 GCAGTATTCC ATTTTGAAG CAGCCGCTTT ACCGGACAAA TCAGCGGCGG  
 2551 CAgGATACG GCATTACACT TAAAGACAG CGAATGGACG CTGCCGTCAG  
 20 2601 GarCGGAATT AGGCAATTTA AACCTTGACA ACGCCACCAT TACaCTCAAT  
 2651 TCCGCCTATC GCCACGATGC GGCAGGGGCG CAAACCGGCA GTGCGACAGA  
 2701 TGCCCGCGCG CGCCGTTCGC GCCGTTCGCG CCGTTCCTTA TTATmCGTTA  
 2751 CAGCGCCAAC TTCGTTAGAA TCCCGTTTCA ACACGCTGAC GGTAAACGGC  
 2801 AAATTGAACG GTCAGGGAAC ATTCCGCTTT ATGTCGGAAC TCTTCGGCTA  
 25 2851 CCGCAGCGAC AAATTGAAGC TGGCGGAAAG TTCCGAAGGC ACTTACACCT  
 2901 TGGCGGTCAA CAATACCGGC AACGAACCTG CAAGCCTCGA ACAATTGACG  
 2951 GTAGTGAAG GAAAAGACAA CAAACCGCTG TCCGAAACC TTAATTTAC  
 3001 CCTGCAAAAC GAACACGTCG ATGCAGGCGC GTGG.....  
 //  
 30 3551 ..... TTAGAC CGCGTATTG CCGAAGACCG  
 3601 CCGCAACGCC GTTTGGACAA GCGGCATCCG GGACACCAA CACTACCGTT  
 3651 CGCAAGATTT CCGCGCTAC CGCAACAAA CCGACCTGCG CCAATCGGT  
 3701 ATGCAGAAAA ACCTCGGCAG CGGGCGCGTC GGCATCCTGT TTTCGCACAA  
 35 3751 CCGGACCGAA AACACCTTCG ACGACGGCAT CGGCAACTCG GCACGGCTTG  
 3801 CCCACGGCGC CGTTTTCGGG CAATACGGCA TCGACAGGTT CTACATCGGC  
 3851 ATCAG<sub>2</sub>CGCG GCGCGGGTT TTAGCAGCGG CAGCCTTTcA GACGGCATCG  
 3901 GAGsmAAAwT CCGCCGCCGC GTGctGCATT ACGGCATTCA GGCACGAtAC  
 3951 CGCGCCGgtt tCggCGgAtT CCGCATCGAA CCGCACATCG GCGCAACGcg  
 4001 ctATTTCGTC CAAAAGCGG ATTACCGCTA CGAAAACGTC AATATCGCCA  
 40 4051 CCCC CGCCT TGCATTCAAC CGcTACCGCG CGGGCATTaa GGCAGATTAT  
 4101 TCATTCAAAC CGGCGCAACA CATTTCATC ACGCCTTATT TGAGCCTGTC  
 4151 CTATACCGAT GCCGCTTCGG GCAAAGTCG AACACGCGTC AATACCGCCG  
 4201 TATTGGCTCA GGATTTCGGC AAAACCCGCA GTGCGGAATG GGgCGTAAAC  
 4251 GCCGAAATCA AAGGTTTCAC GCTGTCCCTC CACGCTGCCG CCGCCAAAGG  
 45 4301 CCGCAACTG GAAGCGCAAC ACAGCGCGGG CATCAAATTA GGCTACCGCT  
 4351 GGTAA...

This corresponds to the amino acid sequence <SEQ ID 648; ORF1>:

1 MKTTDKRTE THRKA<sub>1</sub>PTGR IRFXAAYLAI CLSFGILPQA WAGHTYFGIN  
 50 51 YQYYRDAEN KGKFAVGA<sub>2</sub>KD IEVYNKKGEL VGKSMTKAPM IDFSVVS<sub>3</sub>RNG  
 101 VAALVG<sub>4</sub>VQYI VSAHNGGYN NVDFGAEGXN IXDQXR<sub>5</sub>TYK IVKRNNYKAG  
 151 TRGHPYGGDY HMPRLH<sub>6</sub>KXVT DAE<sub>7</sub>PVEMTSY MDGRKYIDQ<sub>8</sub>N NYPDRVRIGA  
 201 GRQY<sub>9</sub>WRSDED EPN<sub>10</sub>NRESSYH IAS.....GS PMFIYDAQKQ  
 251 KWLINGVLQT GNPYIGKSNG FQLVRK<sub>11</sub>DW<sub>12</sub>FY DEIFAGDTHS VFYEPRQNGK  
 301 YSFND<sub>13</sub>DNNGT GKINAKHEHN SLPNRLK<sub>14</sub>TRT VQLEFNVSLSE TAREPVYHAA  
 351 GGVNSYR<sub>15</sub>PRL NNGENISFID EGK<sub>16</sub>GELILTS NINQAGGLY FQGDFTVSPE  
 401 NNETWQ<sub>17</sub>GAGV HISEDSTVTW KVN<sub>18</sub>GVANDRL SKIGK<sub>19</sub>TL...  
 //  
 60 701 .....DKVTAS LTKTDISGNV DLADHAHLNL TGLATLNGNL  
 751 SANGDTRYTV SHNATQNGNX SLVXNAQATF NQATLNGNTS ASGNASFNLS  
 801 DHAVQNGSLT LSGNAKANVS H<sub>20</sub>SALNGNVSL ADKAVFHFES SRETGQISGG  
 851 KDTALHLKDS EWTLP<sub>21</sub>SGXEL GNLNL<sub>22</sub>DNATI TLNSAYRHDA AGAQTGSATD  
 901 APRRRSR<sub>23</sub>RSR RSL<sub>24</sub>LXVTPPT SVESR<sub>25</sub>ENTLT VNGKLNGQGT FREMSELFY  
 951 RSDKLKLAES SEGTYTLAVN NTGNEPASLE QLT<sub>26</sub>VVEGKDN KPLSENLNFT  
 1001 LQNEHVDAGA W.....  
 //  
 65 1151 .....LDRVFAEDR  
 1201 RNAVWTSGIR DTKHYRSQDF RAYRQQTDLR QIGMQKNLGS GRVGILFSHN  
 1251 RTENTFDDGI GNSARLAHGA VFGQY<sub>27</sub>GIDRF YIGISAGAGF SSGSLSDGIG  
 1301 KXKRRRVLHY GIQARYRAGF GFGIEPHIG ATRYFVQKAD YRYENVN<sub>28</sub>IAT  
 70 1351 PGLAFNRYRA GIKADYSFKP AQHISITPYL SLSYTDAA<sub>29</sub>SG KVRTRVNTAV

1401 LAQDFGKTRS AEWGVNAEIK GFTLSLHAAA AKGPQLEAQH SAGIKLGYRW  
 1451 \*

Further sequencing analysis revealed the complete nucleotide sequence <SEQ ID 649>:

1 ATGAAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCGAA  
 5 51 AACCGGCCGC ATCCGCTTCT CGCCTGCTTA CTTAGCCATA TGCCTGTCTG  
 101 TCGGCATTCT TCCCCAAGCC TGGGCGGGAC ACACTTATTT CGGCATCAAC  
 151 TACCAATACT ATCGCGACTT TGCCGAAAT AAAGGCAAGT TTGCAGTCGG  
 201 GCGCAAGAT ATTGAGGTTT ACAACAAAA AGGGGAGTTG GTCGGCAAAT  
 251 CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTG GCGTAACGGC  
 10 301 GTGGCGGCAT TGGTGGGCGA TCAATATATT GTGAGCGTGG CACATAACGG  
 351 CGGCTATAAC AACGTTGATT TTGGTGCGGA AGGAAGAAAT CCCGATCAAC  
 401 ATCGTTTAC TTATAAAATT GTGAAACGGA ATAATTATAA AGCAGGGACT  
 451 AAAGGCCATC CTTATGGCGG CGATTATCAT ATGCCCGGTT TGCATAAATT  
 501 TGACACAGAT GCAGAACCTG TTGAAATGAC CAGTTATATG GATGGGCGGA  
 15 551 AATATATCGA TCAAAATAT TACCCTGACC GTGTTCTGAT TGGGGCAGGC  
 601 AGGCAATATT GCGCATCTGA TGAAGATGAG CCCAATAACC GCGAAAGTTC  
 651 ATATCATATT GCAAGTGGCT ATTCTTGGCT CGTTGGTGGC AATACCTTTG  
 701 CACAAATGG ATCAGGTGGT GGCACAGTCA ACTTAGGTAG TGAAAAAATT  
 20 751 AAACATAGCC CATATGGTTT TTTACCAACA GGAGGCTCAT TTGGCGACAG  
 801 TGGCTACCA ATGTTTATCT ATGATGCCCA AAAGCAAAAG TGGTTAATTA  
 851 ATGGGGTATT GCAAACGGGC AACCCCTATA TAGGAAAAAG CAATGGCTTC  
 901 CAGCTGGTTC GTAAAGATTG GTTCTATGAT GAAATCTTTG CTGGAGATAC  
 951 CCATTAGTA TTCTACGAAC CACGTCAAAA TGGGAAATAC TCTTTTAACG  
 1001 ACGATAATAA TGGCACAGGA AAAATCAATG CCAAACATGA ACACAATTCT  
 25 1051 CTGCCTAATA GATTAAAAAC ACGAACCGTT CAATTGTTTA ATGTTTCTTT  
 1101 ATCCGAGACA GCAAGAGAAC CTGTTTATCA TGCTGCAGGT GGTGTCAACA  
 1151 TTATCGACC CAGACTGAAT AATGGAGAAA ATATTTCTTT TATTGACGAA  
 1201 GGAAAAGCGC AATTGATACT TACCAGCAAC ATCAATCAAG GTGCTGGAGG  
 30 1251 ATTATATTTC CAAGGAGATT TTACGGTCTC GCCTGAAAT AACGAACTT  
 1301 GGCAAGGCGC GGGCGTTTAT ATCAGTGAAG ACAGTACCGT TACTTGGAAA  
 1351 GTAAACGGCG TGGCAAACGA CCGCCTGTCC AAAATCGGCA AAGGCACGCT  
 1401 GCACGTTCAA GCCAAAGGGG AAAACCAAGG CTCGATCAGC GTGGGCGACG  
 1451 GTACAGTCAT TTTGGATCAG CAGGCAGACG ATAAAGGCAA AAAACAGCC  
 35 1501 TTTAGTGAAG TCGGCTTGGT CAGCGCAGG GGTACGGTGC AACTGAATGC  
 1551 CGATAATCAG TTCAACCCCG ACAAACTCTA TTTCGGCTTT CGCGGCGGAC  
 1601 GTTTGGATTT AAACGGGCAT TCGCTTTCGT TCCACCGTAT TCAAAATACC  
 1651 GATGAAGGGG CGATGATTGT CAACCACAAT CAAGACAAAG AATCCACCGT  
 1701 TACCATTACA GGCAATAAAG ATATTGCTAC AACCGGCAAT AACAACAGCT  
 40 1751 TGGATAGCAA AAAAGAAATT GCCTACAACG GTTGGTTTGG CGAGAAAGAT  
 1801 ACGACCAAAA CGAACGGGCG GCTCAACCTT GTTACCAGC CCGCCGAGA  
 1851 AGACCGCACC CTGCTGCTTT CCGCGGGAAC AAATTTAAAC GGCAACATCA  
 1901 CGCAACAAA CGGCAAACTG TTTTTCAGCG GCAGACCAAC ACCGCACGCC  
 1951 TACAATCATT TAAACGACCA TTGGTTCGCA AAAGAGGGCA TTCCTCGCGG  
 45 2001 GGAATCGTG TGGGACAACG ACTGGATCAA CCGCACATTT AAAGCGGAAA  
 2051 AACTTCAAAT TAAAGCGGCA CAGGCGGTGG TTTCCGCAA TGTTCGCAA  
 2101 GTGAAAGCGC ATTGGCATTT GAGCAATCAC GCCCAAGCAG TTTTGTGTG  
 2151 CGCACCGCAT CAAAGCCACA CAATCTGTAC ACGTTCGGAC TGGACGGGTC  
 2201 TGACAAATTG TGTCGAAAAA ACCATTACCG ACGATAAAGT GATTGCTTCA  
 50 2251 TTGACTAAGA CCGACATCAG CGGCAATGTC GATCTTCCG ATCAGCTCA  
 2301 TTTAAATCTC ACAGGGCTTG CCACACTCAA CGGCAATCTT AGTGCAATG  
 2351 GCGATACAGC TTATACAGTC AGCCACAACG CCACCAAAA CGGCAACCTT  
 2401 AGCCTCGTGG GCAATGCCCA AGCAACATTT AATCAAGCCA CATTAAACGG  
 2451 CAACACATCG GCTTCGGGCA ATGCTTCATT TAATCTAAGC GACCACGCCG  
 55 2501 TACAAAACGG CAGTCTGACG CTTTCCGGCA ACGCTAAGGC AAACGTAAGC  
 2551 CATTCCGCAC TCAACGGTAA TGTCTCCCTA GCCGATAAGG CAGTATTCCA  
 2601 TTTTGAAAGC AGCCGCTTTA CCGGACAAAT CAGCGGCGGC AAGGATACGG  
 2651 CATTACACTT AAAAGACAGC GAATGGACGC TGCCGTCAGG CACGGAATTA  
 2701 GGCAATTTAA ACCTTGACAA CGCCACCATT ACACTCAATT CCGCCTATCG  
 2751 CCACGATGCG GCAGGGGCGC AAACCGGCAG TCGGACAGAT GCGCCGCGCC  
 60 2801 GCGTTCGCG CCGTTCGCGC CGTTCCTAT TATCCGTTAC ACCGCCAACT  
 2851 TCGGTAGAAT CCCGTTTCAA CACGCTGACG GTAAACGGCA AATTGAACGG  
 2901 TCAGGAACA TTCCGCTTTA TGTGGAACCT CTTCCGCTAC CGCAGCGACA  
 2951 AATTGAAGCT GCGGAAAAGT TCCGAAGGCA CTTACACCTT GCGGTC AAC  
 3001 AATACCGGCA ACGAACCTGC AAGCCTCGAA CAATTGACGG TAGTGAAGG  
 65 3051 AAAAGACAAC AAACCGCTGT CCGAAACCT TAATTTACC CTGCAAAACG  
 3101 AACACGTCGA TGCCGGCGCG TGGCGTTACC AACTCATCCG CAAAGACGGC  
 3151 GAGTTCGCC TGCAATATCC GGTCAAAGAA CAAGAGCTTT CCGACAACT  
 3201 CGGCAAGGCA GAAGCCAAAA AACAGCGGGA AAAAGACAAC GCGCAAGGCC  
 3251 TTGACGCGCT GATTGCGGCC GGGCGGATG CCGTCGAAAA GACAGAAAGC  
 70 3301 GTTGCCGAAC CGGCCGCGCA GGCAGCGGGG GAAATGTGCG GCATTATGCA

-361-

5  
10  
15  
20

```

3351 GGC GGAGGAA GAGAAAAAAC GGGTGCAGGC GGATAAAGAC ACCGCCTTGG
3401 CGAAACAGCG CGAAGCGGAA ACCCGGCCGG CTACCACCGC CTCCCCCGC
3451 CCCGCCGCG CCCGCCGGA TTGCGCGCAA CTGCAACCCC AACCGCAGCC
3501 CCAACCGCAG CGCGACCTGA TCAGCCGTTA TGCCAATAGC GGT TTGAGTG
3551 AATTTTCCGC CACGCTCAAC AGCGTTTTTCG CCGTACAGGA CGAATTAGAC
3601 CGCGTATTTG CCGAAGACCG CCGCAACGCC GTTTGGACAA GCGGCATCCG
3651 GGACACCAA CACTACCGTT CGCAAGATT CCGCGCCTAC CGCCAACAAA
3701 CCGACCTGCG CCAAATCGGT ATGCAGAAAA ACCTCGGCAG CCGGCGCGTC
3751 GGCATCCTGT TTTCGCACAA CCGGACCGAA AACACCTTCG ACGACGGCAT
3801 CCGCAACTCG GCACGGCTTG CCCACGGCGC CGTTTTTCGG CAATACGGCA
3851 TCGACAGGTT CTACATCGGC ATCAGCGCGG GCGCGGGTTT TAGCAGCGGC
3901 AGCCTTTTCAG ACGGCATCGG AGGCAAAATC CGCGCGCGCG TGCTGCATTA
3951 CCGCATTCAG GCACGATACC GCGCGGGTTT CCGCGGATTC GGCATCGAAC
4001 CGCACATCGG CGCAACGCGC TATTTTCGTC AAAAAGCGGA TTACCGCTAC
4051 GAAACGTCAT ATATCGCCAC CCCCGGCCTT GCATTCAACC GCTACCGCGC
4101 GGCCTTAAG GCACGATATT CATTCAAACC GGCGCAACAC ATTTCATCA
4151 CGCCTTATTT GAGCCTGTCC TATACCGATG CCGCTTCGGG CAAAGTCCGA
4201 ACACGCGTCA ATACCGCCGT ATTGGCTCAG GATTTTCGCA AAACCGCAG
4251 TCGGGAATGG GCGGTAAACG CCGAAATCAA AGGTTTCACG CTGTCCCTCC
4301 ACGTGCCGC CGCCAAGGC CCGCAACTGG AAGCGCAACA CAGCGCGGGC
4351 ATCAATTAG GCTACCGCTG GTAA

```

This corresponds to the amino acid sequence <SEQ ID 650; ORF1-1>:

25  
30  
35  
40  
45  
50

```

1 MKTTDKRTTE THRKAPKTGR IREFSPAYLAI CLSFGILPQA WAGHTYFGIN
51 YQYYRDFEEN KGFVAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVS RNG
101 VAALVGDQYI VSAHNGGYN NVDFGAEGRN PDQHRFTYKI VKRNNYKAGT
151 KGHYPYGGDYH MPRLHKFVTD AEPVEMTSYM DGRKYIDQNN YPDRVRIGAG
201 RQYWRSEDEE PNNRESSYHI ASAYSWLVGG NTFAQNGSGG GTVNLGSEKI
251 KHSPYGFLEP GGSFGDSGSP MFIYDAQKQK WLINGVLQTG NPYIGKSNFG
301 QLVRKDWFYD EIFAGDTHSV FYEPRQNGKY SFNDDNNGTG KINAKHEHNS
351 LPNRLKTRTV QLENVSLSET AREPVYHAAG GVNSYRPRLN NGENISFIDE
401 KGKELILTSN INQAGGLYF QGDFTVSPEN NETWQAGVH ISEDSTVTWK
451 VNGVANDRLS KIGKGTLHVQ AKGENQGSIS VGDGTVILDQ QADDKGGKQA
501 FSEIGLVSGR GTVQLNADNQ FNPDKLYFGF RGGRLDLNGH SLSFHRIQNT
551 DEGAMIVNHN QDKESTVTIT GNKDIATTGN NNSLDSKKEI AYNWFGEKD
601 TTKTNGLRNL VYQPAEDRT LLLSGGTNLN GNITQTNGKL FFSGRPTPHA
651 YNHLNDHWSQ KEGIPRGEIV WDNDWINRTF KAENFQIKGG QAVVSRNVAK
701 VKGDWHLNSH AQAVFGVAPH QSHTICTRSD WTGLTNCVEK TITDDKVIAS
751 LTKTDISGNV DLADHAHLNL TGLATLNGNL SANGDTRYTV SHNATQNGNL
801 SLVGNAQATF NQATLNGNTS ASGNASFNLS DHAVQNGSLT LSGNAKANVS
851 HSALNGNVSL ADKAVFHFES SRFTGQISGG KDTALHLKDS EWTLPSTEL
901 GNLNLDNATI TLNSAYRHDA AGAQTGSATD APRRRSRRSR RSLLSVTPPT
951 SVESRFTLT VNGKLNQGT FRFMSELFY RSDKLKLAES SEGTYTLAVN
1001 NTGNEPASLE QLTVEGKDN KPLSENLFNT LQNEHVDAGA WRYQLIRKDG
1051 EFRLHNPVKE QELSDKLGA EAKKQAEKDN AQSLDALIAA GRDAVEKTES
1101 VAEPARQAGG ENVGIMQAE EKKRVQADKD TALAKQREAE TRPATTAFFR
1151 ARRARRDLPO LQPQPQPQPQ RDLISRYANS GLSEFSATLN SVFAVQDELD
1201 RVFAEDRRNA VWTSGIRDTK HYRSQDFRAY RQQTDLRQIG MQKNLGSGRV
1251 GILFSHNRT E NTDDGIGNS ARLAHGAVFG QYGIDRFYIG ISAGAGFSSG
1301 SLSDGIGGKI RRRVLHYIQ ARYRAGFGGF GIEPHIGATR YFVQKADYRY
1351 ENVNIATPGL AFNRYRAGIK ADYSFKPAQH ISITPYLSLS YTDAASGKVR
1401 TRVNTAVLAQ DFGKTRSAEW GVNAEIKGFT LSLHAAAAGK PQLEAQHSAG
1451 IKLGYRW*

```

Computer analysis of these sequences gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

55 ORF1 shows 57.8% identity over a 1456aa overlap with an ORF (ORF1a) from strain A of *N. meningitidis*:

60

```

          10      20      30      40      50      60
orfl.pep MKTTDKRTTETHR KAPKTGR IREFXAYLAICLSFGILPQAWAGHTYFGIN YQYYRDFEEN
          |||||
orfla    MKTTDKRTTETHR KAPKTGR IREFSPAYLAICLSFGILPQAWAGHTYFGIN YQYYRDFEEN
          10      20      30      40      50      60

          70      80      90     100     110     120
orfl.pep KGFVAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVS RNGVAALVGVQYIVSAHNGGYN

```

-362-

orfla  
70 80 90 100 110 120  
5  
orfl.pep 130 140 150 160 170 180  
orfla  
10  
orfl.pep 190 200 210  
orfla  
15  
orfl.pep 220 230 240 250 260  
orfla  
20  
orfl.pep 270 280 290 300 310 320  
orfla  
25  
orfl.pep 330 340 350 360 370 380  
orfla  
30  
orfl.pep 390 400 410 420 430  
orfla  
35  
orfl.pep  
orfla  
40  
orfl.pep  
orfla  
45  
orfl.pep  
orfla  
50  
orfl.pep  
orfla  
55  
orfl.pep  
orfla  
60  
orfl.pep  
orfla  
65  
orfl.pep  
orfla  
70  
orfl.pep

KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVSRNGVAALVGDQYIVSVAHNGGYN  
NVDFGAEGXNIXDQXRXYKIVKRNNYKAGTKGHPYGGDYHMPRLHKXVTD AEPVEMTSY  
NVDFGAEGXN-PDQHRFSYQIVKRNNYKPDNS-HPYNGDXHMPRLHKFVTD AEPVEMTSY  
MDGRKYIDQNNYPDRVRIGAGRQYWRSD EDEP-----NN-----  
MRGNTYS DKEKYP E RVRIGSGHHYWR YDDDKHGDLSYSGAWLIGGNTHMQGWGNNGVXSL  
SGDVRHANDYGPMPIAGAAGDSGSPMFIYDKTNNKWLNGVLQGTYPYSGRENGFQLIRK  
DWFYDEIFAGDTHSVFYEP RQNGKYSFNDDNNGTGKINAKHEHNSLPNRLKTRTVOLFNV  
DWFYDDIYRGDTHTVXFEPRSNHGFSTSNNGTGTVTETNEKVSNP-KLKVQTVRLFDE  
SLSETAREPVYHAAGGVNSYRPRLNNGENISFIDEKGELILTSNINQAGGLYFQGDFT  
SLNETDKEPVY-AAGGVNQYRPRLNNGENLSFIDYGNGLILSNINQAGGLYFEGDFT  
VSPENNETWQAGGVHISEDSTVTWKVNGVANDRLSKIGKGTL-----  
VSPENNETWQAGGVHISEDSTVTWKVNGVANDRLSKIGKGT LHVQAKGENQGSISVGDGT  
VILDQQADDKGGKQAFSEIGLXSGRGT VQLNADNQFNPDKLYFGFRGRLDLNGHSLSFH  
RIQNTDEGAMIXXHNATTTSTVTITGNESITQPSGKNINRLNYSKEIAYNGWFGEKD TTK  
TNGRLNLVYQPAEDRTXLLSGGTNLNGNITQTNGKLFFSGRPTPHAYNHLGSGWSKMEG  
IPQGEIVWDNDWIXRTFKAENFHIQGGQAVISRNVAKVEGDHLSNHAQAVFGVAPHQSH  
-----XXXXXDKVTASLTKTDISGNVDLADHAHLNLTGLATLNGNLSAN  
TICTRSDWTGLTNCVEXXITDDKVIASLTKTDXSGXVXLXXXXXXXXLXGXAXLXGNLSAN  
GDTRYTVSHNATQNGNXSLVXNAQATFNQATLNGNTSASGNASFNLSDAH VQNGSITLSG

70 The complete length ORF1a nucleotide sequence <SEQ ID 651> is:



5

10

15

20

25

30

35

40

45

50

55

60

65

70

1 ATGAAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCGAA  
51 AACCGGCCGC ATCCGCTTCT CGCCTGCTTA CTTAGCCATA TGCCTGTCGT  
101 TCCGGATTCT TCCCCAAGCT TGGGCGGGAC ACACCTATTT CGGCATCAAC  
151 TACCAATACT ATCGCGACTT TGCCGAAAAT AAAGGCAAGT TTGCAGTCGG  
201 GCGCAAGAT ATTGAGGTNT ACAACAAAA AGGGGAGTTG GTCGGCAAAT  
251 CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTG GCGTAACGGC  
301 GTGGCGGCAT TGGTGGCGGA TCAATATATT GTGAGCGTGG CACATAACGG  
351 CCGCTATAAC AACGTTGATT TTGGTGCGGA AGGAAGNAAT CCCGATCAGC  
401 ACCGTTTTTC TTACCAAAAT GTGAAAAGAA ATAATTATAA GCCTGACAAT  
451 TCACACCCTT ACAACGGCGA TTANCATATG CCGCGTTTGC ATAAATTTGT  
501 CACAGATGCA GAACCTGTCT AAATGACGAG TGACATGAGG GGGAAATACCT  
551 ATTCGATATA AGAAAAATAT CCCGAGCGTG TCCGCATCGG CTCAGGACAC  
601 CACTATTGGC GTTATGATGA TGACAAACAC GCGGATTAT CCTACTCCGG  
651 CGCATGGTTA ATTGGCGGCA ATACACATAT GCAGGGTTGG GGAAATAATG  
701 GCGTANTTAG TTTGAGCGGC GATGTGCGCC ATGCCAACGA CTATGGCCCT  
751 ATGCCGATTG CAGGTGCGGC AGGCGACAGC GGTTCCGCCA TGTTTATTTA  
801 TGACAAAACA AACAATAAAT GGCTGCTCAA CCGAGTTTTA CAAACCGGT  
851 ACCCTTATTC CCGCAGGGAA AACGGTTTCC AGCTGATACG CAAAGATTGG  
901 TTCTACGATG ACATTTACAG AGGCGATACA CATACCGTCT NTTTTGAACC  
951 CGCGAGTAAC GGACATTTTT CCTTTACATC CAACAACAAC GGTACGGGTA  
1001 CGGTAACAGA AACCAACGAA AAGGTNTCCA ATCCAAAGCT TAAAGTACAG  
1051 ACAGTCCGAC TGTTTGACGA ATCTTTGAAT GAAACTGATA AAGAACCAGT  
1101 TTACCGCGCA GGGGGTGTTA ATCAGTACCG TCCAAGGTTA AACACCGGTG  
1151 AAAACCTTTC TTTTATCGAT TACGCAACG GCAAACTCAT CTTATCAAAC  
1201 AACATCAACC AAGCGCGGG CGGTTGTAT TTTGAAGGTG ATTTTACGGT  
1251 CTCGCTGAA AACACGAAA CGTGCAAGG CCGGGCGCTT CATATCAGTG  
1301 AAGACAGTAC CGTTACTTGG AAAGTAAACG GCGTGGCAAA CGACCGCCTG  
1351 TCCAAAATCG GCAAAGGCAC GCTGCACGTT CAAGCCAAAG GGGAAAACCA  
1401 AGGTCGATC AGCGTGGGCG ACGGTACAGT CATTTTGGAT CAGCAGGCAG  
1451 ACGATAAAGG CAAAAACAA GCCTTTAGTG AAATCGGCTT GNTCAGCGCG  
1501 AGGGGTACGG TGCAACTGAA TGCCGATAAT CAGTTCAACC CCGACAAACT  
1551 CTATTTCCGGC TTTCGCGCGG GACGTTTGGA TTAAACGGG CATTGCTTTT  
1601 CGTTCACCG TATTCAAAAT ACCGATGAAG GGGCGATGAT TGNCNATCAT  
1651 AATGCCACAA CAACATCCAC CGTTACCATT ACAGGGAATG AAAGTATTAC  
1701 ACAACCGAGT GGTAAGAATA TCAATAGACT TAATTACAGC AAAGAAATTG  
1751 CCTACAACGG TTGGTTTGGC GAGAAAGATA CGACCAAAAC GAACGGGCGG  
1801 CTCAACCTTG TTTACCAGCC CGCCGCAGAA GACCGCACCC NGCTGCTTTC  
1851 CCGCGGAACA AATTAAACG GCAACATCAC GCAAACAAAC GGCAAACTGT  
1901 TTTTCAGCGG CAGACCGACA CCGCACGCCT ACAATCATTT AGGAAGCGGG  
1951 TGGTCAAAAA TGGAAGGTAT CCCACAAGGA GAAATCGTGT GGGACAACGA  
2001 CTGGATCNAC CGCACGTTTA AAGCGGAAAA TTTCCATATT CAGGGCGGGC  
2051 AGGCGGTGAT TTCCCGCAAT GTTGCCAAAG TGGAAGGCGA TTGNCATTG  
2101 AGCAATCACG CCCAAGCAGT TTTTGGTGTG GCACCGCATC AAAGCCATAC  
2151 AATCTGTACA CGTTCGACT GGACNGGTCT GACAAATTGT GTCGAANAAA  
2201 NCATTACCGA CGATAAAGTG ATTGCTTCAT TGACTAAGAC NGACNTNAGC  
2251 GGCANTGTNA GNCTNNCCNA TNACGNTNNT TNAAANCTCN CNGGGCNTGC  
2301 NNCACTNAAN GGCAATCTTA GTGCAATGG CGATACAGT TATACAGTCA  
2351 GCCACAACGC CACCCAAAAC GGCAACCTTA GCCTCGTGGG CAATGCCCAA  
2401 GCAACATTTA ATCAAGGCAC ATTAACGGC AACNCATCGG NTTCCGGCAA  
2451 TGCTTCATTT AATCTAAGCA ACAACGCCG ACAAAACGGC AGTCTGACGC  
2501 TTTCCGACAA CGCTAAGGCA AACGTAAGCC ATTCCGCACT CAACGGCAAT  
2551 GTCTCCCTAG CCGATAAGGC AGTATTCCAT TTTGAAAACA GCCGCTTTAC  
2601 CGGACAACCT AGCGGCAGCA AGGANACAGC ATTACACTTA AAAGACAGCG  
2651 AATGGACGCT GCCGTACGGC ACGGAATTAG GCAATTTAAA CCTTGACAAC  
2701 GCCACCATTA CACTCAATTC CGCCTATCGC CACGATGCTG CAGGCGCGCA  
2751 AACCGGCAGN GTGTACAGCA CGCCGCGCCG CCGTTCGCGC CGTTCCTAT  
2801 TATCCGTTAC ACCGCCAATC TCGGTAGAAT CCCGTTTCAA CACGCTGACG  
2851 GTAAACGGCA AATTGAACNG TCAAGGAACA TTCCGCTTTA TGTCGGAAT  
2901 CTTCCGCTAC CGAAGCGACA AATTGAAGCT GGCGGAAAGT TCCGAAGGNA  
2951 CTTACACCTT GCGGTCAAC AATACCGGCA ACGAACCCGT AAGCCTCGAT  
3001 CAATTGACGG TAGTGGAAAG GAAAGACAAC AAACCGCTGT CCGAAAACCT  
3051 TAATTTACC CTGCAAAACG AACACGTCGA TGCCGGCGCG TGCGGTACC  
3101 AACTCATCCG CAAAGACGGC GAGTTCGGCC TGCATATCC GGTCAAAGAA  
3151 CAAGAGCTTT CCGACAACT CGGCAAGGCA GAAGCCAAAA AACAGGCGGA  
3201 AAAAGACAAC GCGCAAGCC TTGACGCGCT GATTGCGGCC GGGCGCGAT  
3251 CCGCCGAAAA GACAGAAAGC GTTGCCGAAC CGGCCCGGCN GGCAGGCGGG  
3301 GAAAATGTCG GCATTATGCA GCGGAGGAA GAGAAAAAAC GGGTGCAGGC  
3351 GGATAAAGAC AGCGCNTTGG CGAAACAGCG CGAAGCGGAA ACCCGGCCGG  
3401 NTACCACCGC CTTCCNCCGC GCCCGCNGCG CCCGCCGGGA TTTGCGCAA  
3451 CCGCAGCCCC AACCGCAACC TCAACCCCAA CCGCAGCGCG ACCTGATNAG  
3501 CCGTTATGCC AATAGCGGTT TGAGTGAATT TTCCGCCACG CTCAACAGCG  
3551 TTTTCGCCGT ACAGGACGAA TTGACCGCG TGTTTGCCGA AGACCGCCGC

-365-

5  
 10  
 15

```

3601 AACGCNGTTT GGACAAGCNG CATCCGGNAC ACCAAACACT ACCGTTTCGCA
3651 AGATTTCGCG GCCTACCGCC AACAAACCGA CCTGCGCCAA ATCGGTATGC
3701 AGAAAAACCT CGGCAGCGGG CGCGTCGGCA TCCTGTTTTC GCACAAACGG
3751 ACCGAAAACA NCTTCGACGA CGGCATCGGC AACTCGGCAC GGCTTGCCCA
3801 CCGCGCCGTT TTCGGGCAAT ACGGCATCGG CAGGTTTCGAC ATCGGCATCA
3851 GCACGGGCGC GGGTTTTAGC AGCGGCANTC TNTCAGACGG CATCGGAGGC
3901 AAAATCCGCC GCCGCGTGCT GCATTACGGC ATTCAGGCAC GATACCGCGC
3951 CCGTTTCGGC GGATTTCGGCA TCGAACCCTA CATCGGCGCA ACGCGCTATT
4001 TCGTCCAAAA AGCGGATTAC CGCTACGAAA ACGTCAATAT CGCCACCCCC
4051 GGTCTTTCGT TCAACCGNTA CCGNGCGGGC ATTAAGGCAG ATTATTCATT
4101 CAAACCGGCG CAACACATNT CCATCACNCC TTATTTNAGC CTGTCCTATA
4151 CCGATGCCCG TTCCGGGCAA GTCCGAACAC GCGTCAATAC CGCNGTATTG
4201 GCTCAGGATT TCGGCAAAAC CCGCAGTGGC GAATGGGGCG TAAACGCCGA
4251 AATCAAAGGT TTCACGCTGT CCNTCCACGC TGCCGCGCGC AAAGNCCGCG
4301 AACTGGAAGC GCAACACAGC GCGGGCATCA AATTAGGCTA CCGCTGGTAA
  
```

This encodes a protein having amino acid sequence <SEQ ID 652>:

20  
 25  
 30  
 35  
 40  
 45

```

1 MKTTDKRTTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA WAGHTYFGIN
51 YQYYRDFAEK KGKFAVGAKD IEVYNKKGEL VGKSMKAPM IDFSVSRNG
101 VAALVGDQYI VSAHNGGYN NVDFGAEGXN PDQHRFSYQI VKRNNYKPDN
151 SHPYNGDXHM PRLHKFVTD AEPVEMTSDMR GNTYSDKEYI PERVRIGSGH
201 HYWRYDDDKH GDLSSGAWL IGGNTHMQGW GNNGVXSLSG DVRHANDYGP
251 MPIAGAAGDS GSPMFIYDKT NNKWLNLGVL QTGYPSGRE NGFQLIRKDW
301 FYDDIYRGDT HTVXFEPKSN GHFSFTSNNN GTGTVTETNE KVSNNPKLVQ
351 TVRLFDESIN ETDKEPVYAA GGVNQYRPRL NNGENLSFID YGNGKLILSN
401 NINQAGGLY FEGDFTVSPE NNETWQAGV HISEDSTVW KVNVDANDRL
451 SKIGKGLHV QAKENQSGI SVGDGTVID QOADDKGGKQ AFSEIGLXSG
501 RGTVQLNADN QFNPKLYFG FRGGRDLNG HSLSFHRIQ TDEGAMIXXH
551 NATTTSTVTI TGNEISITQP GKNINRLNYS KEIAYNGWFG EKDTTKTNGR
601 LNLVQPAAE DRTXLLSGT NLNGNITQTN GKLFSSGRPT PHAYNHLGSG
651 WSKMEGIPQG EIVWDDWIX RTFKAENFHI QGGQAVISR N VAKVEGDHXL
701 SNHAQAVFGV APHQSHTICT RSDWTGLTNC VEXXITDDKV IASLTKTDXS
751 GXVXLXXXXX XXLGXAXLX GNLSANGDTR YTVSHNATON GNLSLVGNAQ
801 ATFNQATLNG NXSXSGNASF NLSNNAQNG SLTSLDNAK NVSHSALNGN
851 VSLADKAVFH FENSRTGQL SGSKXTALHL KDSEWTLPSG TELGNLNDN
901 ATITLNSAYR HDAAGAQTGX VSDTPRRRSR RSLLSVTPPT SVESRENTLT
951 VNGKLNQGT FRFMSEIFGY RSDKLKLAES SEGTYTLAVN NTGNEPVSLD
1001 QLTVEGKDN KPLSENLFNT LQNEHVDAGA WRYQLIRKDG EFRHLNVPKE
1051 QELSDKLGA EAKKQAEKDN AQSLDALIAA GRDAEKTES VAEPARXAGG
1101 ENVGIMQAE EKKRVQADKD SALAKQREAE TRPXTTAFPR ARXARRDLPO
1151 PQQPQPQPQ PQRDLXSRYA NSGLSEFSAT LNSVFAVQDE LDRVFAEDRR
1201 NAVWTSKIRX TKHYRSQDFR AYRQQTDLRQ IGMQKNLGS RVGILFSHNR
1251 TENXFDGIG NSARLAHGA FGQYIGIRFD IGISTGAGFS SGXLSDGIGG
1301 KIRRRVLHYG IQARYRAGFG GFGIEPYIGA TRYFVQKADY RYENVNIATP
1351 GLAFNRYRAG IKADYSFKPA QHXSITPYXS LSYTDAASGK VRTRVNTAVL
1401 AQDFGKTRSA EWGVNAEIKG FTLSXHAAAA KGPQLEAQS AGIKLGYRW*
  
```

A transmembrane region is underlined.

ORF1-1 shows 86.3% identity over a 1462aa overlap with ORF1a:

50  
 55  
 60  
 65

```

      10      20      30      40      50      60
orfla.pep MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINYQYYRDFAEK
      |||
orfl-1    MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINYQYYRDFAEK
      10      20      30      40      50      60

      70      80      90     100     110     120
orfla.pep KGKFAVGAKDIEVYNKKGELVGKSMKAPMIDFSVSRNGVAALVGDQYIVSAHNGGYN
      |||
orfl-1    KGKFAVGAKDIEVYNKKGELVGKSMKAPMIDFSVSRNGVAALVGDQYIVSAHNGGYN
      70      80      90     100     110     120

      130     140     150     160     170     179
orfla.pep NVDFGAEGXNPDQHRFSYQIVKRNNYKPDNS-HPYNGDXHMPRLHKFVTD AEPVEMTSDM
      |||
orfl-1    NVDFGAEGRNPDQHRFTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKFVTD AEPVEMTSYM
      130     140     150     160     170     180
  
```

10

15

20

25

30

35

40

45

50

60

65

70

-367-

		900	910	920	930	940
	orfla.pep	TELGNLNLDNATITLNSAYRHDAAGAQ	TGXVSDTPRRRSRRS---	LLSVTPPTS	SVESRFN	
	orfl-1	TELGNLNLDNATITLNSAYRHDAAGAQ	TGSATDAPRRRSRRSRRS	LLSVTPPTS	SVESRFN	
5		900	910	920	930	940
	orfla.pep	950	960	970	980	990
	orfl-1	950	960	970	980	990
10		1000	1010	1020	1030	1040
	orfla.pep	TLTVNGKLNQGTFRFMSE	LFGYRSDKLKLAESSE	GYTLAVNNTGNEPVSLD	QLTVVEG	
	orfl-1	TLTVNGKLNQGTFRFMSE	LFGYRSDKLKLAESSE	GYTLAVNNTGNEPASLE	QLTVVEG	
		960	970	980	990	1000
	orfla.pep	1010	1020	1030	1040	1050
	orfl-1	1010	1020	1030	1040	1050
15		1060	1070	1080	1090	1100
	orfla.pep	KDNKPLSENLNFTLQNEHVDAGAWRYQLIRKDG	EFRLHNPVKEQELSDKL	GKAEAKKQAE		
	orfl-1	KDNKPLSENLNFTLQNEHVDAGAWRYQLIRKDG	EFRLHNPVKEQELSDKL	GKAEAKKQAE		
		1020	1030	1040	1050	1060
	orfla.pep	1070	1080	1090	1100	1110
	orfl-1	1070	1080	1090	1100	1110
20		1120	1130	1140	1150	1160
	orfla.pep	KDNAQSLDALIAAGRDAAEKTESVAEPARXAGGENVGIMQAE	EEKKRVQADKDSALAKQR			
	orfl-1	KDNAQSLDALIAAGRDAVEKTESVAEPARQAGGENVGIMQAE	EEKKRVQADKDTALAKQR			
		1080	1090	1100	1110	1120
	orfla.pep	1130	1140	1150	1160	1170
	orfl-1	1130	1140	1150	1160	1170
25		1180	1190	1200	1210	1220
	orfla.pep	EAETRPXTTAFPRARXARRDL	PQPQPQPQPQORDLX	SRYANSGLSEFSATLNSVFAV		
	orfl-1	EAETRPATTAFPRARRARRDL	PQLQPQPQPQ--QRDL	ISRYANSGLSEFSATLNSVFAV		
		1140	1150	1160	1170	1180
	orfla.pep	1190	1200	1210	1220	1230
	orfl-1	1190	1200	1210	1220	1230
30		1240	1250	1260	1270	1280
	orfla.pep	QDELDRVFAEDRRNAVWTSXIRXTKHYRSQDFRAYRQOTDLRQIGM	QKNLGS	SGRVGILFS		
	orfl-1	QDELDRVFAEDRRNAVWTSXIRXTKHYRSQDFRAYRQOTDLRQIGM	QKNLGS	SGRVGILFS		
		1200	1210	1220	1230	1240
	orfla.pep	1250	1260	1270	1280	1290
	orfl-1	1250	1260	1270	1280	1290
35		1300	1310	1320	1330	1340
	orfla.pep	HNRTENXFDDGIGNSARLAHGA	VFGQYIGRFDIGISTGAGFSSG	XLSDGIGGKIRRRVL		
	orfl-1	HNRTENTFDDGIGNSARLAHGA	VFGQYIDRFYIGISAGAGFSSG	SLSDGIGGKIRRRVL		
		1260	1270	1280	1290	1300
	orfla.pep	1310	1320	1330	1340	1350
	orfl-1	1310	1320	1330	1340	1350
40		1360	1370	1380	1390	1400
	orfla.pep	HYGIQARYRAGFGGFGIEPYIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSF				
	orfl-1	HYGIQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSF				
		1320	1330	1340	1350	1360
	orfla.pep	1370	1380	1390	1400	1410
	orfl-1	1370	1380	1390	1400	1410
45		1420	1430	1440	1450	1460
	orfla.pep	KPAQHXSITPYXLSYTDAA	SGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSXHA			
	orfl-1	KPAQHISITPYLSYTDAA	SGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHA			
		1380	1390	1400	1410	1420
	orfla.pep	1430	1440	1450	1460	1470
	orfl-1	1430	1440	1450	1460	1470
50		1480	1490	1500	1510	1520
	orfla.pep	AAAKGPQLEAQHSAGIKLGYR	WX			
	orfl-1	AAAKGPQLEAQHSAGIKLGYR	WX			
		1440	1450	1460	1470	1480

Homology with adhesion and penetration protein hap precursor of *H.influenzae* (accession number P45387)

Amino acids 23-423 of ORF1 show 59% aa identity with hap protein in 450aa overlap:

orfl	23	FXAAYLAICLSFGILPQAWAGHTYFGIN	YQYYR	DF	AE	NGKGF	AVGAKDIEVYNKKGELVG	82
hap	6	FRLNFLTACVSLGIASQAWAGHTYFGID	YQYYR	DF	AE	NGKGF	TVGAKNIEVYNKEGQLVG	65
orfl	83	KSMTKAPMIDFSVVS	RNGVAALVG	QYIV	SAHNGGYN	VD	FGAEGXNIXDQXRXYKIV	142
hap	66	TSMTKAPMIDFSVVS	RNGVAALVG	QYIV	SAHNGGYN	VD	FGAEG N DQ R TY+IV	124

5  
10  
15  
20  
25

orf1 143 KRNNYKAGTKGHPYGGDYHMPRLHKXVTDAEPVEMTSYMDGRKYIDQNNYPDRVRIGAGR 202  
KRNNY+A + HPY GDYHMPRLHK VT+AEFV MT+ MDG+ Y D+ NYP+RVRIG+GR  
hap 125 KRNNYQAWERKHPYDGDYHMPRLHKFVTEAEFVGMTTNMDGKVYADRENYPERVRIGSGR 184

orf1 203 QYWRSEDEPNRESSYHIA----- 222  
QYWR+D+DE N SSY+++  
hap 185 QYWRTDKDEETNVHSSYYVSGAYRYLTAGNTHTQSGNGNGTVNLSGNVVSPPNHYGPLPTG 244

orf1 223 -----SGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNGFQLVRKDWFYDEIFAGDTHSVF 277  
SGSPMFIYDA+K++WLIN VLQTG+P+ G+ NGFQL+R++WFY+E+ A DT SVF  
hap 245 GSKGDSGSPMFIYDAKKKQWLINAVLQTGHPFFGRGNGFQLIREEFYNEVLAVDTPSVF 304

orf1 278 --YEPRONGKYSFNDDNNGTGKIN-AKHEHNSLPNRLKTRTVQLFNVSLSETAREPVYHA 334  
Y P NG YSF +N+GTGK+ + + + + TV+LFN SL++TA+E V A  
hap 305 QRYIPPINGHYSFVSNNDGTGKLTLRPSKDGSKAKSEVGTVKLFNPSLNQTAKHEV-KA 363

orf1 335 AGGVNSYRPRLNNGENISFIDEKGELILTSNINQAGGLYFQGDFTV-SPENNETWQGA 393  
A G N Y+PR+ G+NI D+GKG L + +NINQAGGLYF+G+F V +NN TWQGA  
hap 364 AAGYNIYQPRMEYGNKIYLGQKGKTLTIENNINQAGGLYFEGNFVVKGKQNNITWQGA 423

orf1 394 GVHISEDSTVTWKVNGVANDRLSKIGKGT 423  
GV I +D+TV WKV+ NDRLSKIG GTL  
hap 424 GVSIGQDATVEWKVHNPENDRLSKIGIGT 453

Amino acids 715-1011 of ORF1 show 50% aa identity with hap protein in 258aa overlap:

30  
35  
40  
45

Orf1 41 DTRYTVSHNATQ-NGNXSLVXNAQATFNQ-ATLNGNTSASGNASFNLSDHAVQNGSLTLS 98  
DT+ S TQ NG+ +L NA + A LNGN + ++ F LS++A Q G++ LS  
hap 733 DTKVINSIPITQINGSINLTNNATVNIHGLAKLNGNVTLIDHSQFTLSNNATQTGNIKLS 792

orf1 99 GNAKANVSHSALNGNVSLADKAVHFHESRFTGQISGGKDTALHLKDSEWTLPSGXELGN 158  
+A A V+++ LNGNV L D A F ++S F QI G KDT + L+++ WT+PS L N  
hap 793 NHANATVNNATLNGNVHLTDSAQFSLKNSHFWHQIQGDKDTTVTLENATWTMPSDTTLQN 852

orf1 159 LNLDNATITLNSAYRHAAGAQTGSATDAPXXXXXXXXXXXXLXVTPPTSVESRENTLTVN 218  
L L+N+T+TLNSAY + S+ +AP L T PTS E RENTLTVN  
hap 853 LTLNNSTVTLNSAY-----SASSNNAPRHRS-----LETETTPSAEHRENTLTVN 899

orf1 219 GKLNQGGTFRFMSELEFYRSCLKLAESSEGTYYTLAVNNTGNEPASLEQLTVVEGKDNKP 278  
GKL+GQGTFF+ S LFGY+SDKLKL+ +EG YTL+V NTG EP +LEQLT++E DNKP  
hap 900 GKLSGQGTFFQFTSSLEFYKSDKLKLSNDAEGDYTLVSRNTGKEPVTLEQLTLIESLDNKP 959

orf1 279 LSENLFNFTLQNEHVDAGA 296  
LS+ L FTL+N+HVDAGA  
hap 960 LSDKLKFTLENDHVDAGA 977

Amino acids 1192-1450 of ORF1 show 41% aa identity with hap protein in 259aa overlap:

50  
55  
60  
65

Orf1 1 LDRVFAEDRRNAVWTSIGIRDTKHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGLFSHNR 60  
LDR+F + ++AVWT+ +D + Y S FRAY+Q+T+LRQIG+QK L +GR+G +FSH+R  
hap 1135 LDRLFVDQAQSAVWNTNIAQDKRRYDSDAFRAYQQTNLRLQIGVQKALANGRIGAVFSHR 1194

orf1 61 TENTFDDGIGNSARLAHGAVFQGYGIDRFYXXXXXXXXXXXXXXXXXIGXKRRRVLHYG 120  
++NTFD+ + N A L + F QY K R+ ++YG  
hap 1195 SDNTFDEQVKNHATLTMMSGFAQYQWGDLOFGVNVGTGISASKMAEEQSRKIHRKAINYG 1254

orf1 121 IQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPA 180  
+ A Y+ G GI+P+ G RYF+++ +Y+ E V + TP LAFNRY AGI+ DY+F P  
hap 1255 VNASYQFRLGQLGIQPYFGVNRYFIERENYQSEEVVRKTPSLAFNRYNAGIRVDYTFPT 1314

orf1 181 QHISITPYLSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAAA 240  
+IS+ PY ++Y D ++ V+T VN VL Q FG+ E G+ AEI F +S + +  
hap 1315 DNISVKPYFFVNYVDVSNANVQTTVNLTVLQQPFGRYWQKEVGLKAEILHFQISAFISKS 1374

orf1 241 KGPQLEAQHSAGIKLGYRW 259  
+G QL Q + G+KLGYRW  
hap 1375 QGSQGLGKQNVGVKLGYSRW 1393

Homology with a predicted ORF from *N.gonorrhoeae*

The blocks of ORF1 show 83.5%, 88.3%, and 97.7% identities in 467, 298, and 259 aa overlap, respectively with a predicted ORF (ORF1ng) from *N.gonorrhoeae*:

5	orf1.pep	MKTTDKRTTETHRKAPKTGRIRFXAAYLAICLSFGILPQAWAGHTYFGINYQYYRDAEN	60
	orf1ng	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINYQYYRDAEN	60
10	orf1.pep	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVS RNGVAALVG VQYIVSVAHNGGYN	120
	orf1ng	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVS RNGVAALAGDQYIVSVAHNGGYN	120
15	orf1.pep	NVDFGAEGXNIXDQXRXYKIVKRNNYKAGTKGHPYGGDYHMPRLHKXVTDAEPVEMTSY	180
	orf1ng	NVDFGAEGSN-PDQHRFSYQIVKRNNYKAGTNGHPYGGDYHMPRLHKFVTDAEPVEMTSY	179
20	orf1.pep	MDGRKYIDQNNYPDRVRIGAGRQYWRSDDEPNRESSYHIAS-----	223
	orf1ng	MDGWKYADLNKYPDRVRIGAGRQYWRSDDEPNRESSYHIASAYSWLVGGNTFAQNGSG	239
25	orf1.pep	-----GSPMFIYDAQKQWLINGVLQTGNFYIGKSNG	255
	orf1ng	GGTVNLGSEKIKHSPYGLPTGGSGSGSPMFIYDAQKQWLINGVLQTGNFYIGKSNG	289
30	orf1.pep	FOLVRKDWFYDEIFAGDTHSVFYEPQNGKYSFNDNNGTGKINAKHEHNSLPNRLKTRT	315
	orf1ng	FOLVRKDWFYDEIFAGDTHSVFYEPHONGKYFFNDNNGAGKIDAKHKHYSLPYRLKTRT	359
35	orf1.pep	VQLENVSLSETAREPVYHAAGGVNSYRPLNNGENISFIDEGKGELILTSNINQGAGGLY	375
	orf1ng	VQLENVSLSETAREPVYHAAGGVNSYRPLNNGENISFIDKGKGELILTSNINQGAGGLY	
40	orf1.pep	FQGDFTVSPENNETWQAGVHISEDSTVTWKVNGVANDRLSKIGKGT	422
	orf1ng	FEGNFTVSPKNNETWQAGVHISDGSTVTWKVNGVANDRLSKIGKGTLLVQAKGENQGSV	479
45	orf1.pep	// DKVTASLTKTDISGNVDLADHAHLNLTGLA	744
	orf1ng	FGVAPHQSHTICTRSDWTGLTSCTEKITDDKVIASLSKTDVRGNVSLADHAHLNLTGLA	774
50	orf1.pep	TLNGNLSANGDTR-YTVSHNATQNGXSLVXNAQATFNQATLNGNTSASGNASFNLSNHA	803
	orf1ng	TFNGNL-VQAEIRTLRANATQNGNLSLVGNAQATFNQATLNGNTSASDNASFNLSNNA	833
55	orf1.pep	VQNGSLTSLGNAKANVSHSALNGNVSLADKAVHFESSRFTGQISGGKDTALHLKDSEWT	863
	orf1ng	VQNGSLTSLDNKANVSHSALNGNVSLADKAVHFENSRTGKISGGKDTALHLKDSEWT	893
60	orf1.pep	LPSGXELGNLNDNATITLNSAYRHAAGAQTGSATDAPRRRSRRSRLXVTPPTSVE	923
	orf1ng	LPSGTTELGNLNDNATITLNSAYRHAAGAQTGSAADAPRRRSRRS---LLSVTPPTSVE	950
65	orf1.pep	SRFNTLTVNGKLNQGTFRFMSELFYRSDKLKLAESSEGYTTLAVNNTGNEPASLEQLT	983
	orf1ng	SRFNTLTVNGKLNQGTFRFMSELFYRSGKLKLAESSEGYTTLAVNNTGNEPVSLEQLT	1010
70	orf1.pep	VVEGKDNKPLSENLFNLQNEHVDAGAW	1011
	orf1ng	VVEGKDNTPLSENLFNLQNEHVDAGAWRYQLIRKDGFEFLHNPVKEQELSDKLKGAGET	1070
75	orf1.pep	// LDRVFAEDRRNAVWTSIGIRDTKHYRSQDFR	1211
	orf1ng	PQRDLISRYANSGLSEFSATLNSVFAVQDELDRVFAEDRRNAVWTSIGIRDTKHYRSQDFR	1239
80	orf1.pep	AYRQQTDLRQIGMQKNLGSGRVGILFSHNRTENTFDDGIGNSARLAHGA VFGQYIGIDRFY	1271
	orf1ng	AYRQQTDLRQIGMQKNLGSGRVGILFSHNRTGNTFDDGIGNSARLAHGA VFGQYIGIRFD	1299

```

orfl.pep      IGISAGAGFSSGSLSDGIGKXRRRLVHYGIQARYRAGFGGFGIEPHIGATRYFVQKADY 1331
|||||
orflng        IGISAGAGFSSGSLSDGIRGKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADY 1359
5
orfl.pep      RYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTTDAASGKVRTRVNTAVL 1391
|||||
orflng        RYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTTDAASGKVRTRVNTAVL 1419
10
orfl.pep      AQDFGKTRSAEWGVNAEIKGFTLSLHAAAAGKPQLEAQHSAGIKLGYRW 1440
|||||
orflng        AQDFGKTRSAEWGVNAEIKGFTLSLHAAAAGKPQLEAQHSAGIKLGYRW 1468

```

The complete length ORF1ng nucleotide sequence was identified <SEQ ID 653>:

```

1  ATGAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCTAA
15 51  AACCCGCCG ATCCGCTTCT CGCCCGCTTA CTTAGCCATA TGCCTGTGCT
101 TCGGCATTCT GCCCAAGGCC CGGGCGGGAC ACACTTATTT CGGCATCAAC
151 TACCAATACT ATCGCGACTT TGCCGAAAAT AAAGGCAAGT TTGCAGTCGG
201 GCGGAAAGAT ATTGAGGTTT ACAACAAAAA AGGGGAGTTG GTCGGCAAAT
251 CGATGACGAA AGCCCCGATG ATTGATTTTT CTGTGGTATC GCGTAACGGC
301 GTGGCGGCAT TGGCGGGCGA TCAATATATT GTGAGCGTGG CACATAACGG
20 351 CGGCTATAAC AATGTTGATT TTGGTGCGGA GGAAGCAAT CCCGATCAGC
401 ACCGCTTTTC TTACCAAAAT GTGAAAAGAA ATAATTATAA AGCAGGGACT
451 AACGGCCATC CTTATGGCGG CGATTATCAT ATGCCGCGTT TGCACAAATT
501 TGTCACAGAT GCAGAACCTG TTGAGATGAC CAGTTATATG GATGGGTGGA
25 551 AATACGCTGA TTAAATAAAA TACCCTGATC GTGTTCAAT CGGAGCAGGC
601 AGACAATATT GCGGCTCTGA TGAAGACGAA CCCAATAACC GCGAAAGTTC
651 ATATCATATT GCAAGCGCAT ATTCTTGGCT CGTCGGTGGC AATACCTTTG
701 CACAAAATGG ATCAGGTGGT GGCACAGTCA ACTTAGGTAG CGAAAAAATT
751 AAACATAGCC CATATGGTTT TTTACCAACA GGAGGCTCAT TTGGCGACAG
30 801 TGGCTCACCA ATGTTTATCT ATGATGCCCC AAAGCAAAAG TGGTTAATTA
851 ATGGGGTATT GCAAACAGGC AACCCTATA TAGGAAAAG CAATGGCTTC
901 CAGCTAGTTC GTAAAGATTG GTTCTATGAT GAAATCTTTG CTGGAGATAC
951 CCATTAGTA TTCTACGAAC CACATCAAAA TGGGAAATAC TTTTTTAACG
1001 ACAATAATAA TGGCGCAGGA AAAATCGATG CCAACATAA ACACATTCT
35 1051 CTACCTTATA GATTAATAAC ACGAACCGTT CAATTGTTTA ATGTTTCTTT
1101 ATCCGAGACA GCAAGAGAAC CTGTTTATCA TGCTGCAGGT GGGGTCAACA
1151 GTTATCGACC CAGACTGAAT AATGGAGAAA ATATTCCTT TATTGACAAA
1201 GGAAGAGGTG AATTGATACT TACCAGCAAC ATCAACCAAG GCGCGGGCGG
1251 TTTGTATTTT GAGGGTAATT TTACGGTCTC GCCTAAAAAC AACGAAACGT
40 1301 GGCAAGGCGC GGGCGTTTAT ATCAGTGATG GCAGTACCGT TACTTGAAAA
1351 GTAAACGGCG TGGCAAACGA CCGCTGTGCC AAAATCGGCA AAGGCACGCT
1401 GCTGTTTCAA GCCAAAGGGG AAAACCAAGG CTCGGTCAGC GTGGGCGAGC
1451 GTAAAGTCAT CTTAGATCAG CAGGCGGACG ATCAAGGCAA AAAACAAGCC
1501 TTTAGTGAAG TCGGCTTGGT CAGCGGCAGG GGGACGGTGC AACTGAATGC
45 1551 CGATAATCAG TTCAACCCCG ACAAACCTTA TTTCGGCTTT CGCGGCGGAC
1601 GTTTGGATTT GAACGGGCAT TCGCTTTCGT TCCACCGCAT TCAAAATACC
1651 GATGAAGGGG CGATGATTGT CAACCACAAT CAAGACAAAG AATCCACCGT
1701 TACCATTACA GGCAATAAAG ATATTACTAC AACCGGCAAT AACACAACCT
1751 TGGATAGCAA AAAAGAAATT GCCTACAACG GTTGGTTTGG CGAGAAAGAT
50 1801 GCAACCAAAA CGAACGGGCG GCTCAATCTG AATTACCAAC CGGAAGAAGC
1851 GGATCGCACT TTACTGCTTT CCGGCGGAAC AAATTTAAAC GGCAATATCA
1901 CGCAACAAA CGGCAAACTG TTTTTCAGCG GCAGACCGAC ACCGCACGCC
1951 TACAATCATT TAGGAAGCGG GTGGTCAAAA ATGGAAGGTA TCCACAAGG
2001 AGAAATCGTG TGGGACAACG ATTGGATCGA CCGCACATT AAAGCGGAAA
2051 ACTTCCATAT TCAGGGCGGA CAAGCGGTGG TTTCCGCAA TGTGCCAAA
55 2101 GTGAAGGCG ATTGGCATT AAGCAATCAC GCCAAGCAG TTTTCGGTGT
2151 CGCACCGCAT CAAAGCCACA CAATCTGTAC ACGTTCGGAC TGGACGGGTC
2201 TGACAAGTTG TACCGAAAAA ACCATTACCG ACGATAAAGT GATTGCTTCA
2251 TTGAGCAAGA CCGACATCAG AGGCAATGTC AGCCTTGCCG ATCAGGCTCA
60 2301 TTTAAATCTC ACAGGACTTG CCACACTCAA CGGCAATCT AGTGAGGCG
2351 GAGACACGCA CTATACGGTT ACGCGCAACG CCACCAAAA CGGCAACCTC
2401 AGCCTCGTGG GCAATGCCCA AGCAACATTT AATCAAGCCA CATTAAACGG
2451 CAACACATCG GCTTCGGACA ATGCTTCATT TAATCTAAGC AACACGCGG
2501 TACAAAACGG CAGTCTGACG CTTTCCGACA ACGCTAAGGC AAACGTAAGC
65 2551 CATTCCGCAC TCAACGGCAA TGTCTCCCTA GCCGATAAGG CAGTATTCCA
2601 TTTTGAAAAC AGCCGCTTTA CCGGAAAAAT CAGCGGCGGC AAGGATACGG
2651 CATTACACTT AAAAGACAGC GAATGGACGC TGCCGTCGGG CACGGAATTA
2701 GGCAATTTAA ACCTTGACAA CGCCACCATT AACTCAATT CCGCCTATCG
2751 ACACGATGCG GCAGGCGCGC AAACCGGCAG TGCGGCAGAT GCGCGCGCC
2801 GCGGTCGCG CCGTTCCTTA TTATCCGTTA CGCCGCCAAC TTCGGCAGAA
70 2851 TCCGTTTCA ACACGCTGAC GGTAAACGGC AAATTGAACG GTCAGGGAAC

```

-371-

2901 ATTCCGCTTT ATGTCGGAAC TCTTCGGCTA CCGCAGCGGC AAATTGAAGC  
 2951 TGGCGGAAAG TTCCGAAGGC ACTTACACCT TGGCTGTCAA CAATACCGGC  
 3001 AACGAACCCG TAAGTCTCGA GCAATTGACG GTAGTGGAAG GAAAAGACAA  
 3051 CACACCGCTG TCCGAAAATC TTAATTTTAC CCTGCaaaAc gaacacgtcg  
 5 3101 atgccggcgca atggCGTTAT CAGCTTATCC gcaaagacgG CGAGTTCCgC  
 3151 CTGCATAATC CGGTCAAAGA ACAAGAGCTT TCCGACAAAC TCGGCAAGgc  
 3201 gggagaaaACA GAggcgcctT TGACGGCAAA ACAGGCacaA CTTGCCGCCA  
 3251 AAcaacaggc ggaAAAAGAC AACgcgcaaa gccttgAcgc gctgattgcg  
 10 3301 gCggggcgca atgccaccga AAAGGCAGaa agtggtgccc aaccgGCCCCG  
 3351 GCAGGCAGGC GGGGAAAAtg ccgGCATTAT GCAGGCGGAG GAAGAGAAAA  
 3401 AACGGGTGCA GCGCGATAAA GACACCGCCT TGGCGAAACA GCGCGAAGCG  
 3451 GAAACCCGGC CGGCTACCAC CGCCTTCCCC CGCGCCCGCC GCGCCCGCCG  
 3501 GGATTGTCCG CAACCGCAGC CCAACCGCA ACCCAACCG CAGCGCGACC  
 3551 TGATCAGCCG TTATGCCAAT AGCGGTTTGA GTGAATTTTC CGCCACGCTC  
 15 3601 AACAGCGTTT TCGCCGTACA GGACGAATTG GACCGCGTGT TTGCCGAAGA  
 3651 CCGCCGCAAC GCCGTTTGGa CAAGCGGCAT CCGGGACACC AAACACTACC  
 3701 GTTCGCAAGA TTTCCGCGCC TACCGCCAAC AAACCGACCT GCGCCAAATC  
 3751 GGTATGCAGA AAAACCTCGG CAGCGGGCGC GTCGGCATCC TGTTTTCGCA  
 3801 CAACCGGACC GGAACACCT TCGACGACGG CATCGGCAAC TCGGCACGGC  
 20 3851 TTGCCACCGG TGCCGTTTTC GGGCAATACG GCATCGGCAG GTTCGACATC  
 3901 GGCATCAGCG CGGGCGCGGG TTTTAGTAGC GGCAGCCTTT CAGACGGCAT  
 3951 CAGAGGCAAA ATCCGCGGCC GCGTGCTGCA TTACGGCATT CAGGCAAGAT  
 4001 ACCGCGCAGG TTTCCGCGGA TTCGGCATCG AACCGCACAT CGGCGCAACG  
 4051 CGCTATTTTC TCCAAAAAGC GGATTACCGA TACGAAAACG TCAATATCGC  
 25 4101 CACCCCGGGC CTTCGATTCA ACCGCTACCG CGCGGGCATT AAGGCAGATT  
 4151 ATTCATTCAA ACCGGCGCAA CACATTTCCA TCACGCCTTA TTTGAGCCTG  
 4201 TCCTATACCG ATGCCGCTTC CGGCAAAGTC CGAACGCGCG TCAATACCGC  
 4251 CGTATTGGCG CAGGATTTTC GCAAACCCG CAGTGCGGAA TGGGGCGTAA  
 4301 ACGCCGAAAT CAAAGGTTTC ACGCTGTCCC TCCACGCTGC CGCCGCCAAG  
 30 4351 GGGCCGCAAT TGAAGCGCA GCACAGCGCG GGCATCAAAT TAGGCTACCG  
 4401 CTGGTAA

This is predicted to encode a protein having amino acid sequence <SEQ ID 654>:

1 MKTTDKRTE THRKAPKTGR IRFSPAYLAI CLSFGILPOA RAGHTYFGIN  
 51 YQYYRDFAEK KGKFAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVSRRG  
 35 101 VAALAGDQYI VSAHNGGYN NVDFGAEGSN PDQHRFSYQI VKRNNYKAGT  
 151 NGHPYGGDYH MPRLHKFVTD AEPVEMTSYM DGWKYADLNK YPDRVRIGAG  
 201 RQYWRSEDEE PNNRESSYHI ASAYSWLVG NTFANQNGSG GTVNLGSEKI  
 251 KHSPYGFLEPT GGSEFGDSGSP MFIYDAQKQK WLVINGVLOTG NPYIGKSNGE  
 301 QLVRKDWFEYD EIFAGDTHSV FYEPHQNGKY FFNDNNNGAG KIDAKHKHYS  
 40 351 LPYRLKTRTV QLFNVSLSET AREPVYHAAG GVNSYRPRLN NGENISFIDK  
 401 KGKELILTSN INQAGGLYF EGNFTVSPKN NETWQAGAVH ISDGSVTWTK  
 451 VNGVANDRLS KIGKGTLLVQ AKGENQGSVS VGDGKVILDQ QADDQKKQA  
 501 FSEIGLVSGR GTVQLNADNQ FNPDKLYFGF RGGRLDLNGH SLSEFHRQNT  
 551 DEGAMIVNHN QDKESTVTIT GNKDITTTGN NNNLDSKKEI AYNGWFEKED  
 601 ATKTNGLNL NYPPEADRT LLLSGGTNLN GNITQTNGL FFSGRPTPHA  
 651 YNHLGSGWSK MEGIPQGEIV WDNDWIDRTF KAENFHIQGG QAVVSRNVAK  
 701 VEGDWHLNSH AQAVFVAPH QSHTICTRSD WTGLTSCTEK TITDDKVIAS  
 751 LSKTDVRGNV SLADHAHLNL TGLATFNGNL VQAETrTIRL RANATQNGNL  
 801 SLVGNAQATF NQATLNGNTS ASDNASFNLS NNAVQNGSLT LSDNAKANVS  
 50 851 HSALNGNVSL ADKAVHFHEN SRFTGKISGG KDTALHLKDS EWTLPSTEL  
 901 GNLNLDNATI TLNSAYRHDA AGAQTGSAAD APRRRSRRSL LSVTPPTSAE  
 951 SRFNTLTVNG KLNGQGTFRF MSELFYGRSG KLKLAESSEG TYTLAVNNTG  
 1001 NEPVSLQELT VVEGKDNTPL SENLNTLQNL EHVDAWARY QLIRKDGFEF  
 1051 LHPVKEQEL SDKLKAGET EAALTAKQAQ LAAKQQAED NAQSLDALIA  
 55 1101 AGRNATEKAE SVAEPARQAG GENAGIMQAE EEKRVQADK DTALAKQREA  
 1151 ETRPATTAFR RARRARRDLP QPQPQPQPQ QRDLSRYAN SGLSEFSATL  
 1201 NSVFAVQDEL DRVFAEDRRN AVWTSGIRD KHYRSQDFRA YRQQTDLRQI  
 1251 GMQKNLGSGR VGILFSHNR NTFTDDGIGN SARLAHGAFF GQYIGRFDI  
 1301 GISAGAGFSS GSLSDGIRGK IRRRVLHYGI QARYRAGFGG FGIEPHIGAT  
 60 1351 RYFVQKADYR YENVNIATPG LAFNRYRAGI KADYSFKPAQ HISITPYLSL  
 1401 SYTDAASGV RTRVNTAVLA QDFGKTRSAE WGVNAEIKGF TSLHAAAAK  
 1451 GPQLEAQHSA GIKLGWRW\*

Underlined and double-underlined sequences represent the active site of a serine protease (trypsin family) and an ATP/GTP-binding site motif A (P-loop).

65 ORF1-1 and ORF1ng show 93.7% identity in 1471 aa overlap:



-372-

5	orf1-1.pep	10 20 30 40 50 60 MKTDDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINYQYYRDAEN 
	orf1ng-1	10 20 30 40 50 60 MKTDDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINYQYYRDAEN 
10	orf1-1.pep	70 80 90 100 110 120 KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVS RNGVAALVG DQYIVSVAHNGGYN 
	orf1ng-1	70 80 90 100 110 120 KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVS RNGVAALAGDQYIVSVAHNGGYN 
15	orf1-1.pep	130 140 150 160 170 180 NVDFGAEGRNP DQHRFTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKFVTDAEPVEMTSYM 
	orf1ng-1	130 140 150 160 170 180 NVDFGAEGSNPDQHRFSYQIVKRNNYKAGTNGHPYGGDYHMPRLHKFVTDAEPVEMTSYM 
20	orf1-1.pep	190 200 210 220 230 240 DGRKYIDQNNYPDRVRIGAGRQYWRSD EDEPNNRESSYHIASAYSWL VGNTFAQNGSGG 
	orf1ng-1	190 200 210 220 230 240 DGWKYADLNKYPDRVRIGAGRQYWRSD EDEPNNRESSYHIASAYSWL VGNTFAQNGSGG 
25	orf1-1.pep	250 260 270 280 290 300 GTVNLGSEKIKHSPYGF LPTGGSFGDSGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNGF 
	orf1ng-1	250 260 270 280 290 300 GTVNLGSEKIKHSPYGF LPTGGSFGDSGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNGF 
30	orf1-1.pep	310 320 330 340 350 360 QLVRKDWFYDEIFAGDTHSVFYEP RONGKYSFNDNNGTGKINAKHEHNSLPNRLKTRTV 
	orf1ng-1	310 320 330 340 350 360 QLVRKDWFYDEIFAGDTHSVFYEP HQNGKYFFNDNNGAGKIDAKHFKHYSLPYRLKTRTV 
35	orf1-1.pep	370 380 390 400 410 420 QLFNVSLSETAREPVYHAAGGVNSYR PRLNNGENISFIDEGKGELILTSNINQGAGGLYF 
	orf1ng-1	370 380 390 400 410 420 QLFNVSLSETAREPVYHAAGGVNSYR PRLNNGENISFIDKGKGELILTSNINQGAGGLYF 
40	orf1-1.pep	430 440 450 460 470 480 QGDFTVSPENNETWQAGGVHISED STVTWKVNGVANDRLSKIGKGLHVQAKGENQGSIS :
	orf1ng-1	430 440 450 460 470 480 EGNFTVSPKNNETWQAGGVHISED STVTWKVNGVANDRLSKIGKGLLVQAKGENQGSVS :
45	orf1-1.pep	490 500 510 520 530 540 VGDGTVILDQQADDKGKQAFSEIGLV SGRGTVQLNADNQFNPKLYFGFRGGRDLNGH 
	orf1ng-1	490 500 510 520 530 540 VGDGKVILDQQADDQGKQAFSEIGLV SGRGTVQLNADNQFNPKLYFGFRGGRDLNGH 
50	orf1-1.pep	550 560 570 580 590 600 SLSFHRIQNTDEGAMIVNHNQDK ESTVTITGNKD IATTGNNNSLDSKKEIAYNGWFG EKD 
	orf1ng-1	550 560 570 580 590 600 SLSFHRIQNTDEGAMIVNHNQDK ESTVTITGNKD ITTTGNNNSLDSKKEIAYNGWFG EKD 
55	orf1-1.pep	610 620 630 640 650 660 TTKTNGRLNLVYQPAEDRTLLL SGGTNLNGNITQTNGKLF FSGRPTPHAYNHLNDHWSQ :
	orf1ng-1	610 620 630 640 650 660 ATKTNGRLNLNYQPEEADRTLLL SGGTNLNGNITQTNGKLF FSGRPTPHAYNHLGSGWSK :
60	orf1-1.pep	670 680 690 700 710 720 KEGIPRGEIVWDNDWINRTFKAEN FQIKGGQAVVS RNVAKVKGDWHL SNHAQAVFGVAPH 
	orf1ng-1	670 680 690 700 710 720 MEGIPQGEIVWDNDWIDRTFKAEN FHIQGGQAVVS RNVAKVEGDWHL SNHAQAVFGVAPH 
65	orf1-1.pep	670 680 690 700 710 720 KEGIPRGEIVWDNDWINRTFKAEN FQIKGGQAVVS RNVAKVKGDWHL SNHAQAVFGVAPH 
	orf1ng-1	670 680 690 700 710 720 MEGIPQGEIVWDNDWIDRTFKAEN FHIQGGQAVVS RNVAKVEGDWHL SNHAQAVFGVAPH 
70	orf1-1.pep	670 680 690 700 710 720 KEGIPRGEIVWDNDWINRTFKAEN FQIKGGQAVVS RNVAKVKGDWHL SNHAQAVFGVAPH 
	orf1ng-1	670 680 690 700 710 720 MEGIPQGEIVWDNDWIDRTFKAEN FHIQGGQAVVS RNVAKVEGDWHL SNHAQAVFGVAPH 

10

15

20

25

30

35

40

45

50

55

65

70

-374-

orfl-1.pep 1430 1440 1450  
 KGFTLSLHAAAAGPQLEAQHSAGIKLGYRWX  
 |||||  
 orflng-1 1440 1450 1460  
 KGFTLSLHAAAAGPQLEAQHSAGIKLGYRWX

In addition, ORF1ng shows 55.7% identity with hap protein (P45387) over a 1455aa overlap:

SCORES Init1: 1104 Initn: 4632 Opt: 2680  
 Smith-Waterman score: 5165; 55.7% identity in 1455 aa overlap

orflng-1.pep 10 20 30 40 50 60  
 MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINYQYYRDAEN  
 p45387 MKKTVFERLNFLTACISLGIVSQAWAGHTYFGIDYQYYRDAEN

orflng-1.pep 70 80 90 100 110 120  
 KGKFAVGAKDIEVYNKKGELVGKSMKAPMIDFSVVSRRNGVAALAGDQYIVSVAHNGGYN  
 p45387 KGKFTVGAQNIKVYNKQQLVGTSMKAPMIDFSVVSRRNGVAALVENQYIVSVAHNVGYT

orflng-1.pep 130 140 150 160 170 180  
 NVDFGAEGSNPDQHRFSYQIVKRNKYKAGTNGHPYGGDYHMPRLHKFVTDAEPVEMTSYM  
 p45387 DVDFGAEGNPDQHRFTYKIVKRNKYKD-NLHPYEDDYHNPRLHKFVTEAAPIDMTSNM

orflng-1.pep 190 200 210 220 230 240  
 DGWKYADLNKYPDRVRIGAGRQYWRSEDEPNNRESSYHIASAYSWLVGNTFAQNGSGG  
 p45387 NGSTYSDRTKYPERVRIGSGRQFWNDQDKG-----QVAGAYHYLTAGNTHNQRGAGN

orflng-1.pep 250 260 270 280 290 300  
 GTVNLGSEKIKHSPYGLFTGGSGSGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNGF  
 p45387 GYSYLGSDVRKAGEYGLPLIAGSKGSGSPMFIYDAEKQKWLINGILREGNPFEGKENG

orflng-1.pep 310 320 330 340 350 360  
 QLVKRDWIFYDEIFAGDTHSVFYEPHQNGKYFFNDNNNGAGKIDAKHKHYSLPYRLKTRTV  
 p45387 QLVKRSYF-DEIFERDLHTSLYTRAGNVYITISGNDNGQGSITQKS---GIPSEIK---I

orflng-1.pep 370 380 390 400 410 419  
 QLFNVSLSETAREPVYHAA-GGVNSYRPLNNGENISFIDKKGELILTSNINQGAGGLY  
 p45387 TLANMSLPLKEKDKVHNPRYDGPNIYSPRLNNGETLYFMDQKQSLIFASDINQGAGGLY

orflng-1.pep 420 430 440 450 460 470 479  
 FEGNFTVSPKNNETWQAGVHISDGSTVTWKVNGVANDRLSKIGKGTLLVQAKGENQGSV  
 p45387 FEGNFTVSPNSQTWQAGIHVSENSTVTWKVNGVEHDLRLSKIGKGTLLVQAKGENKGS

orflng-1.pep 480 490 500 510 520 530 539  
 SVGDGKVILDQQAADDQGGKQAFSEIGLVSGRGTVQLNADNQFNPDKLYFGFRGGRDLNLG  
 p45387 SVGDGKVILEQQAADDQGNKQAFSEIGLVSGRGTVQLNDDKQFDTDKFYFGFRGGRDLNLG

orflng-1.pep 540 550 560 570 580 590  
 HSLSFHRIQNTDEGAMIVNHNQDKESTVTITGNKDITT-TGNN-NNLDSKKEIAYNGWFG  
 p45387 HSLTFKRIQNTDEGAMIVNHNNTQAANVTITGNESIVLPNGNNINKLDYRKEIAYNGWFG

5

10

15

20

25

30

35

40

45

50

55

60

70

-376-

		1320	1330	1340	1350	1360	1370
5	orflng-1.pep	SLSDGIRGKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGL					
	p45387	:  : : : : :  :  :  :  : : : : : : :  :  :  : :					
		1240	1250	1260	1270	1280	1290
		KMAEEQSRKIHRKAINYGVNASYQFRLGQLGIQPYFGVNRIFYERENYQSEEVVRVKTPSL					
10	orflng-1.pep	1380	1390	1400	1410	1420	1430
	p45387	AFNRYRAGIKADYSFKPAQHISITPYLSLSYTDAAASGKVRTRVNTAVLAQDFGKTRSAEW					
		:  : : : : :  :  :  :  : : : : : : :  :  :  : :					
		1300	1310	1320	1330	1340	1350
		AFNRYNAGIRVDYTFPTDNISVKPYFFVNYVDVSNANVQTTVNLTVLQQPFGRYQKEV					
15	orflng-1.pep	1440	1450	1460	1469		
	p45387	GVNAEIKGFTLSLHAAAAGKGPQLEAQSAGIKLGYRWX					
		: :  :  :  : :  :  :  : : : : : :					
		1360	1370	1380	1390		
		GLKAEILHFQISAFISKSQGSQGLGKQNVGVKLGYSRW					

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 20 Example 78

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 655>:

	1	..AAGGTGTGGC	AATTGTGCGA	AGA.CCGCTG	CGTGCCGTCG	TGCCTGCCGA
	51	CAGTTTGTAA	CCGACCGCGC	AAAAATTGAA	CCTGTTTAAG	CGCGGTGCGG
25	101	CAACCATTTT	GTTTTATGAA	GATCAAAATG	TCGTCAAAGG	TTTGCAGGAG
	151	CAGTTCCCTG	CTTATGCCCG	TAACTTCCCC	GTTTGGGCGG	ATCAGGCAAA
	201	CGCGATGGTG	CAGTATGCCG	TTTGGACGAC	ACTTGCCGCG	GTCGGCGTAG
	251	GTGCAACCT	GCAACATTAC	AATCCCTTGC	CCGATGCGGC	GATTGCCAAA
	301	GCGTGGAATA	TCCCCGAAAA	CTGGTTGTTG	CGCGCACAAA	TGGTTATCGG
30	351	CGGTATTGAA	GGGCGGCGAG	GTGAAAAGAC	CTTTGAACCC	GTTGCAGAAC
	401	GTTTGAAAGT	GTTGCGCGCA	TAA		

This corresponds to the amino acid sequence <SEQ ID 656; ORF6>:

	1	..KVVQFVEXPL	RAVVPADSFE	PTAQKLNLFK	AGAATILFYE	DQNVVKGLQE
	51	QFPAYAAANF	VWADQANAMV	QYAVWTTLAA	VGVGANLQHY	NPLPDAAIAK
	101	AWNIPENWLL	RAQMVIIGIE	GAAGEKTFEP	VAERLKVFGA	*

35 Further sequence analysis revealed a further partial DNA sequence <SEQ ID 657>:

	1	..CTGCGTGCCG	TCGTGCCTGC	CGACAGTTTT	GAACCGACCG	CGCAAAAATT
	51	GAACCTGTTT	AAGCGGGGTG	CGGCAACCAT	TTTGTTTTAT	GAAGATCAAA
	101	ATGTCGTCAA	AGGTTTGCGAG	GAGCAGTTCC	CTGCTTATGC	CGCTAACTTC
40	151	CCCGTTTGGG	CGGATCAGGC	AAACGCGATG	GTGCAGTATG	CCGTTTGGAC
	201	GACACTTGCC	GCGGTCGGCG	TAGGTGCAAA	CCTGCAACAT	TACAATCCCT
	251	TGCCCCGATG	GCGGATTGCC	AAAGCGTGGA	ATATCCCCGA	AACTGGTTG
	301	TTGCGCGCAC	AAATGGTTAT	CGGCGGTATT	GAAGGGGCGG	CAGGTGAAAA
	351	GACCTTTGAA	CCCGTTGCAG	AACGTTTGAA	AGTGTTCCGC	GCATAA

This corresponds to the amino acid sequence <SEQ ID 658; ORF6-1>:

45	1	..LRVVPADSF	EPTAQKLNLF	KAGAATILFY	EDQNVVKGLQ	EQFPAYAAANF
	51	PVWADQANAM	VQYAVWTTLA	AVGVGANLQH	YNPLPDAAIA	KAWNIPENWL
	101	LRAQMVIGGI	EGAAGEKTFE	PVAERLKVFG	A*	

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

50 ORF6 shows 98.6% identity over a 140aa overlap with an ORF (ORF6a) from strain A of *N.meningitidis*:

-377-

```

      orf6.pep                                KVVQFVEXPLRAVVPADSFEPTAQKLNLFK
      orf6a      QIVEHAVLHTPSSFNSQSARVVVLFGEEDKVVQFVEDALRAVVPADSFEPTAQKLNLFK
                    40      50      60      70      80      90
5
      orf6.pep      AGAATILFYEDQNVVKGLEQFPAYAAANFPVWADQANAMVQYAVWTTLAAVGVGANLQHY
      orf6a      AGAATILFYEDQNVVKGLEQFPAYAAANFPVWADQANAMVQYAVWTTLAAVGVGANLQHY
                    100     110     120     130     140     150
10
      orf6.pep      NPLPDAAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKTFFPVAERLKVFGAX
      orf6a      NPLPDAAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKTFFPVAERLKVFGAX
                    160     170     180     190     200
15

```

The complete length ORF6a nucleotide sequence <SEQ ID 659> is:

```

      1  ATGACCCGTC AATCTCTGCA ACAGGCTGCC GAAAGCCGCC GTTCCATTTA
      51  TTCGTTAAAT AAAAATCTGC CCGTCGGCAA AGATGAAATC GTCCAAATCG
20  101  TCGAACACGC CGTTTTCAC ACACCTTCTT CGTTCAATTC CCAATCTGCC
      151  CGTGTGGTCG TGCTGTTTGG CGAAGAGCAT GATAAGGTGT GGCAATTGT
      201  CGAAGACGCG CTGCGTGCCG TCGTGCCGTC CGACAGTTTT GAACCGACCG
      251  CGCAAAAATT GAACCTGTTT AAGGCGGGTG CGGCAACTAT TTTGTTTAT
      301  GAAGATCAAA ATGTCGTCAA AGGTTTGCAG GAGCAGTTCC CTGCTTATGC
25  351  CGCCAACCTT CCCGTTTGGG CGGACCAGGC GAACGCGATG GTGCAGTATG
      401  CCGTTTGGAC GACACTTGCC GCGGTCGGCG TAGGTGCAAA CCTGCAACAT
      451  TACAATCCCT TGCCCGATGC GCGGATTGCC AAAGCGTGA ATATCCCCGA
      501  AAAGTGGTTG TTGCGCGCAC AAATGGTTAT CGGCGGTATT GAAGGGGCGG
30  551  CAGGTGAAAA GACCTTTGAA CCAGTTGCAG AACGTTTGAA AGTGTTCCGC
      601  GCATAA

```

This is predicted to encode a protein having amino acid sequence <SEQ ID 660>:

```

      1  MTRQSLQQA ESRRSIYSLN KNLVPGKDEI VQIVEHAVLH TPSSFNSQSA
      51  RVVVLFGEEH DKVWQFVEDA LRAVVPADSF EPTAQKLNLF KAGAATILFY
35  101  EDQNVVKGLEQ EQFPAYAAANF PVWADQANAM VQYAVWTTLA AVGVGANLQH
      151  YNPLPDAAIA KAWNIPENWL LRAQMVIIGI EGAAGEKTFF PVAERLKVFG
      201  A*

```

ORF6a and ORF6-1 show 100.0% identity in 131 aa overlap:

```

      40  orf6a.pep      TPSSFNSQSARVVVLFGEEDKVVQFVEDALRAVVPADSFEPTAQKLNLFKAGAATILFY
      orf6-1      LRAVVPADSFEPTAQKLNLFKAGAATILFY
                    50      60      70      80      90     100
                    10      20      30
45  orf6a.pep      EDQNVVKGLEQFPAYAAANFPVWADQANAMVQYAVWTTLAAVGVGANLQHYNPLPDAAIA
      orf6-1      EDQNVVKGLEQFPAYAAANFPVWADQANAMVQYAVWTTLAAVGVGANLQHYNPLPDAAIA
                    110     120     130     140     150     160
                    40      50      60      70      80      90
50  orf6a.pep      KAWNIPENWLLRAQMVIIGGIEGAAGEKTFFPVAERLKVFGAX
      orf6-1      KAWNIPENWLLRAQMVIIGGIEGAAGEKTFFPVAERLKVFGAX
                    170     180     190     200
55  orf6a.pep      KAWNIPENWLLRAQMVIIGGIEGAAGEKTFFPVAERLKVFGAX
      orf6-1      KAWNIPENWLLRAQMVIIGGIEGAAGEKTFFPVAERLKVFGAX
                    100     110     120     130

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF6 shows 95.7% identity over a 140aa overlap with a predicted ORF (ORF6ng) from *N.gonorrhoeae*:

-378-

```

    orf6.pep                                KVVQFVEXPLRAVVPADSFEPTAQKLNLFK  30
    orf6ng                                SNVSLDMSNPTVLRMGLPLYIASLRRGAIYKVVQFVEDALRAVVPADSFEPTAQKLFK  64
5    orf6.pep                                AGAATILFYEDQNVVKGLQEQQFPAYAAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHY  90
    orf6ng                                AGAATILFYEDQNVVKGLQEQQFPAYAAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHY 124
10   orf6.pep                                NPLPDAAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKTFEPVAERLKVFGA  140
    orf6ng                                NPLPDVAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKVFEPAERLKVFGA  174

```

The complete length ORF6ng nucleotide sequence <SEQ ID 661> was identified as:

```

15   1  ATGCCCGTTG CGTCAAATGT CAGCTTGGAT ATGTCCAATC CTACGGTGTG
    51  ACGCATGGGA TTACCCCTTAT ATATTGCGTC CCTAAGAAGG GGCGCAATAT
    101 ATAAGGTGTG GCAATTTGTC GAAGACGCGC TGCCTGCCGT CGTGCCTGCC
    151 GACAGTTTGT AACCGACCGC GCAAAAATTG AAGCTGTTTA AGGCGGGCGC
    201 GGCAACCATT TTGTTTTATG AAGATCAAAA TGTCTCAAA GGTTCGAGG
    251 AGCAGTCCC TGCTTATGCC GCCAACTTTC CCGTTTGGGC GGACCAGGCG
20   301 AACGCTATGG TACAGTATGC CGTCTGGACG ACACTGCGC CGGTCGGTGC
    351 AGGTGCAAA CTGCAACATT ACAACCCCTT GCCCGATGTG GCGATTGCTA
    401 AAGCGTGGAA TATCCCGAA AACTGGCTGT TGC GCGCGCA AATGGTTATC
    451 GGTGGTATG AAGGGGcggc aggtgaaaaa gtctttgaac CCGTTGCgga
    501 acgtttgAAA GTGTTCCGCG CATAA

```

25 This encodes a protein having amino acid sequence <SEQ ID 662>:

```

30   1  MAVASNVSLD MSNPTVLRMG LPLYIASLRR GAIYKVVQFV EDALRAVVP
    51  DSFEPTAQL KLFKAGAATI LFYEDQNVVK GLQEQQPAYA ANFPVWADQ
    101 NAMVQYAVWT TLAAVGAGAN LQHYNPLPDV AIAKAWNIPE NWLLRAQMI
    151 GGIEGAAGEK VFEPVAERLK VFGA*

```

ORF6ng and ORF6-1 show 96.9% identity in 131 aa overlap:

```

35   orf6-1.pep                                10      20      30
    orf6ng                                PTVLRMGLPLYIASLRRGAIYKVVQFVEDALRAVVPADSFEPTAQKLFKAGAAATILFY
    20      30      40      50      60      70
40   orf6-1.pep                                EDQNVVKGLQEQQFPAYAAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHYNPLDAAIA
    orf6ng                                EDQNVVKGLQEQQFPAYAAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHYNPLDVAIA
    80      90      100     110     120     130
45   orf6-1.pep                                100     110     120     130
    orf6ng                                KAWNIPENWLLRAQMVIIGGIEGAAGEKTFEPVAERLKVFGAX
    140     150     160     170

```

50 It is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 79

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 663>

```

55   1  ..GGCTACAAC ACCTGTTGCG GCGCGGCAGC CGCATCGCCA ACTACCAAAT
    51  CAACGGCATC CCCGTTGCCG ACGCGCTGGC CGATACGGG CAATGCCAAC
    101 ACCGCCGCT ATGAGCGCGT AGAAGTCGTG CGCGGCGTGG CGGGGCTGCT
    151 GGACGGCACG GCGGAGCCTT CCGCCACCGT CAATCTGGTG CGCAAACGCC
    201 TGACCCGCAA GCCATTGTTT GAAGTCCGCG CCGAAGCgGG CAACCGcAAA

```

-379-

251 CATTTCGGGC TGGACGCGGA CGTATCGGGC AGCCTGAACA CCGAAG.crC  
 301 rCTGCGCgGC CGCCTGGTTT CCAcCTTCGG ACGCGGCGAC TCGTGGCGGC  
 351 GGCGCGAAGC CAGCCGskAT GCCGAACCTT ACGGCATTTT GGAATACGAC  
 401 ATCGCACCGC AAACCCGCGT CCACGCArGC ATGGACTACC AGCAGGCGAA  
 5 451 AGAAACCGCC GACGCGCCGC TCAGcTACGC CGTGTACGAC AGCCAAGGTT  
 501 ATGCCACCGC CTTCGGCCCG AAAGACAACC CCGCCACAAA TTGGGCGAAC  
 551 AGCCACCACC GTGCGCTCAA CCTGTTCCGC GGCATCGAAC ACCGCTTCAA  
 601 CCAAGACTGG AACTCAAAG CCGAATACGA CTAC..

This corresponds to the amino acid sequence <SEQ ID 664; ORF23>:

10 1 ..GYNLFARGs RIANYQINGI PVADALADTG NANTAAYERV EVVRGVAGLL  
 51 DGTGEPsATV NLVRKRLTRK PLFEVRAEAG NRKHFGLDAD VSGSLNTEXX  
 101 LRGRLVSTFG RGDSWRRRER SRXAELYGIL EYDIAPQTRV HAXMDYQQAk  
 151 ETADAPLSYA VYDSQGYATA FGPKDNPATN WANSHHRLN LFAGIEHREN  
 201 QDWKLKAEYD Y..

15 Further work revealed the complete nucleotide sequence <SEQ ID 665>:

1 ATGACACGCT TCAAATATTC CTTGCTGTTT GCCGCCCTGT TGCCCGTGTA  
 51 CGCGCAGGCC GATGTTTCTG TTTCAGACGA CCCCAAACCG CAGGAAAGCA  
 101 CTGAATTGCC GACCATCACC GTTACCGCCG ACCGCACCGC GAGTTCCTAAC  
 151 GACGGCTACA CTGTTTCCGG CACGCACACC CCGCTCGGGC TGCCCATGAC  
 20 201 CTTGCGCGAA ATCCCGCAGA GCGTCAGCGT CATCACATCG CAACAAATGC  
 251 GCGACCAAAA CATCAAAACG CTCGACCGCG CCCTGTTGCA GCGCACCGGC  
 301 ACCAGCCGCC AGATTTACGG CTCGACCGC GCGGGCTACA ACTACCTGTT  
 351 CGCGCGCGGC AGCCGCATCG CCAACTACCA AATCAACGGC ATCCCCGTTG  
 401 CCGACGCGCT GGCCGATACG GGCAATGCCA ACACCGCCGC CTATGAGCGC  
 25 451 GTAGAAGTCG TGCGCGCGCT GCGGGGGCTG CTGGACGGCA CGGGCGAGCC  
 501 TTCCGCCACC GTCAATCTGG TCGCAAACG CCTGACCGC AAGCCATTGT  
 551 TTGAAGTCGG CGCCGAAGCG GGCAACCGCA AACATTTCCG GCTGGACCGC  
 601 GACGTATCGG GCAGCCTGAA CACCGAAGGC ACGCTGCGCG GCCGCCTGGT  
 651 TTCCACCTTC GGACGCGCGG ACTCGTGGCG GCGGCGCGAA CGCAGCCGCG  
 30 701 ATGCCGAAC CTACGGCATT TTGGAATACG ACATCGCACC GCAAACCCGC  
 751 GTCCACGCAG GCATGGACTA CCAGCAGGCG AAAGAAACCG CCGACGCGCC  
 801 GCTCAGCTAC GCCGTGTACG ACAGCCAAGG TTATGCCACC GCCTTCGGCC  
 851 CGAAAGACAA CCCCGCCACA AATTGGGCGA ACAGCGCCA CCGTGGCTC  
 901 AACCTGTTCG CCGGCATCGA ACACCGCTTC AACCAAGACT GGAAACTCAA  
 35 951 AGCCGAATAC GACTACACCC GCAGCCGCTT CCGCCAGCCC TACGGCGTAG  
 1001 CAGGCGTGCT TTCCATCGAC CACAACACCG CCGCCACCGA CCTGATTCCC  
 1051 GGTATTGGC ACGCCGACCC GCGCACCAC AGCGCCAGCG TGTCATTGAT  
 1101 CGGCAAATAC CGCCTGTTG GCGCGAACA CGATTTAATC GCGGGTATCA  
 1151 ACGGTTACAA ATACGCCAGC AACAAATACG GCGAACGCAG CATCATCCCC  
 40 1201 AACGCCATT CCAACGCCTA CGAATTTTCC CGCACGGGTG CCTACCCGCA  
 1251 GCCTGCATCG TTTGCCAAA CCATCCCGCA ATACGGCACC AGGCGGCAAA  
 1301 TCGGCGGCTA TCTCGCCACC CGTTTCCGCG CCGCGACAA CTTTCCGTG  
 1351 ATTTTGGGCG GACGATACAC CCGTTACCGC ACCGGCAGCT ACGACAGCCG  
 1401 CACACAAGGC ATGACCTATG TGTCCGCCAA CCGTTTACC CCCTACACAG  
 45 1451 GCATCGTGTT CGACCTGACC GGCAACCTGT CTCTTTACGG CTCGTACAGC  
 1501 AGCCTGTTG TCCCGCAATC GCAAAAAGAC GAACACGGCA GCTACCTGAA  
 1551 ACCCGTAACC GGCAACAATC TGGAAGCCGG CATCAAAGGC GAATGGCTTG  
 1601 AAGGCGTCT GAACGCATCC GCCGCCGTGT ACCGCGCCCG TAAAAACAAC  
 1651 CTCGCCACCG CAGCAGGACG CGACCCGAGC GGCAACACCT ACTACCGCGC  
 50 1701 CGCAAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GCGGCGCGCA  
 1751 TCACGCCCGA ATGGCAGATA CAGGCAGGTT ACAGCCAAAG CAAAACCCGC  
 1801 GACCAAGACG GCAGCCGCCT GAACCCGAC AGCGTACCG AACGCAGCTT  
 1851 CAAACTCTTC ACTGCCTACC ACTTTGCCCC CGAAGCCCCC AGCGGTGGA  
 1901 CCATCGGCGC AGGCGTGCGC TGGCAGAGCG AAACCCACAC CGACCTGCC  
 55 1951 ACGTCCGCA TCCCCAACCC CGCCGCCAAA GCCCGCGCCG CCGACACAG  
 2001 CCGCCAAAAA GCCTACGCCG TCGCCGACAT CATGGCGCGT TACCGCTTCA  
 2051 ATCCGCGCGC CGAACTGTG CTGAACGTGG ACAATCTGTT CAACAAACAC  
 2101 TACCGCACCC AGCCCGACCG CCACAGCTAC GCGCACTGC GGACAGTGAA  
 2151 CGCGCGTGT ACCTATCGGT TTAAATAA

60 This corresponds to the amino acid sequence <SEQ ID 666; ORF23-1>:

1 MTRFKYSLLF AALLEPVYAQA DVSVDDEPKP QESTELPTIT VTADRTASSN  
 51 DGYTVSGTHT PLGLEMTLRE IPQSVSVITS QQMRDQNIKT LDRALLQATG  
 101 TSRQIYGSDR AGYNLFARG SRIANYQING IPVADALADT GNANTAAYER  
 151 VEVVRGVAGL LDGTGEPsAT VNLVRKRLTR KPLFEVRAEA GNRKHFGLDA  
 65 201 DVSGSLNTEG TLRGRIVSTF GRGDSWRRRE RSRDAELYGI LEYDIAPQTR



5

251	VHAGMDYQQA	KETADAPLSY	AVYDSQGYAT	AFGPKDNPAT	NWANSRHRAL
301	NLFAGIEHRF	NQDWKLKAEY	DYTRSFRQFP	YGVAGVLSID	HNTAATDLIP
351	GYWHADPRTH	SASVSLIGKY	RLFGREHDLI	AGINGYKYAS	NKYGERSIIP
401	NAIPNAYEFS	RTGAYPOPAS	FAQTIPQYGT	RRQIGGYLAT	RFRADNLSL
451	ILGGRYTRYR	TGSYDSRTQG	MTYVSANRFT	PYTGIVFDLT	GNLSLYGSYS
501	SLFVPQSQKD	EHGSYLKPVT	GNNLEAGIKG	EWLEGRNLAS	AAVYRARKNN
551	LATAAGRDPs	GNTYYRAANQ	AKTHGWEIEV	GGRITPEWQI	QAGYSQSKTR
601	DQDGSRLNPD	SVPERSFKLF	TAYHFAPEAP	SGWTIGAGVR	WQSETHDPA
651	TLRIPNPAK	ARAADNSRQK	AYAVADIMAR	YRFNPRAELS	LNVDNLFNKH
701	YRTOPIRHSY	GALRTVNAAF	TYREF*		

10

**Homology with the ferric-pseudobactin receptor PupB of *Pseudomonas putida* (accession number P38047)**

15	Orf23	6	FARGSRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEP SATVNLVRK	65
	PupB	215	WSRGFAIQNYEVDGVP TSTR L-DNYSQSMAMFDRVEIVRGATGLISGMGNPSATINLIRK	273
20	Orf23	66	RLTRKPLFEVRAEAGNRKHFG LDADVSGSLNTEXXLRGLRVSTFXXXXXXXXXXXXXXAE	125
	PupB	274	R T + + EAGN +G DVSG L +RGR V+ + RPTAEAAQASITGEAGNWD RYGTGF D VSGPLTETGNIRGRFVADYKTEKAWIDRYNQO SQL	333
25	Orf23	126	LYGILEYDIAPQTRVHAXMDYQQAKETADAPLSYAVYD--SQGYATAFGPKDNPATNWAN	183
	PupB	334	+YGI E+D++ T + Y + D+PL + S G T N A +W+ MYGITEFDLSEDTLLTVGFSY--LRSDIDSP LRSGLPTRFSTGERTNLKRS LNAAPDWSY	391
	Orf23	184	SHHRA LNLFAGIEHRFNQDWK LKAE	208
	PupB	392	+ H + F IE + W K E NDHEOTSFFTSIEQQLGNWSGKIE	416

ORF23 shows 95.7% identity over a 211aa overlap with an ORF (ORF23a) from strain A of *N. meningitidis*:

The complete length ORF23a nucleotide sequence <SEO ID 667> is:

-381-

1 ATGACACGCT TCAAATATTC CCTGCTGTTT GCCGCCCTGT TGCCCGTGTA  
 51 CGCGCAGGCC GATGTTTCTG TTTCAGACGA CCCAAAACCG CAGGAAAGCA  
 101 CTGAATTGCC GACCATCACC GTTACCGCGC ACCGCACCGC GAGTTCCAAC  
 151 GACGGCTACA CTGTTTCCGG CACGCACACC CCGCTCGGGC TGCCCATGAC  
 5 201 CCTGCGCGAA ATCCCGCAGA GCGTCAGCGT CATCACATCG CAACAAATGC  
 251 GCGACCAAAA CATCAAAGCG CTCGACCGCG CCCTGTTGCA GCGCAGCCGC  
 301 ACCAGCCGCC AGATTACGG CTCCGACCGC GCGGGCTACA ACTACCTGTT  
 351 CGCGCGCGCG AGCCGCATCG CCAACTACCA AATCAACGGC ATCCCCGTTG  
 401 CCGACGCGCT GGCCGATACG GGCAATGCCA ACACCGCCGC CTATGAGCGC  
 10 451 GTAGAAGTCG TGCGCGGCGT GGCGGGGCTG CTGGACGGCA CGGGCGAGCC  
 501 TTCCGCCACC GTCAATCTGG TGCGCAAACG CCCGACCCGC AAGCCATTGT  
 551 TTGAAGTCGG CGCCGAAGCG GGCAACCGCA AACATTTCGG GCTGGGCGCG  
 601 GACGTATCGG GCAGCCTGAA TGCCGAAGGC ACGCTGCGCG GCCGCTGGT  
 15 651 TTCCACCTTC GGACGCGGCG ACTCGTGGCG GCAGCGCGAA CGCAGCCGCG  
 701 ATGCCGAACCT CTACGGCATT TTGGAATACG ACATCGCACC GCAAAACCCGC  
 751 GTCCACGCAG GCATGGACTA CCAGCAGGCG AAAGAAACCG CCGACGCGCC  
 801 GCTCAGCTAC GCCGTGTACG ACAGCCAAGG TTATGCCACC GCCTTCGGCC  
 851 CGAAAGACAA CCCC GCCACA AATTGGGCGA ACAGCCGCCA CCGTGCCTC  
 901 AACCTGTTGCG CCGGCATCGA ACACCGCTTC AACCAAGACT GGAACCTCAA  
 20 951 AGCCGAATAC GACTACACCC GCAGCCGCTT CCGCCAGCCC TACGGCGTAG  
 1001 CAGGCGTGCT TTCCATCGAC CACAACACCG CCGCCACCGA CTGATTCC  
 1051 GGTTATTGGC ACGCCGACCC GCGCACCCAC AGCGCCAGCG TGTCATTAAT  
 1101 CGGCAAATAC CGCCTGTTG GCGCGAACA CGATTTAATC GCGGGTATCA  
 1151 ACGGTTACAA ATACGCCAGC AACAAATACG GCGAACGCAG CATCATCCCC  
 25 1201 AACGCCATT CCAACGCCTA CGAATTTTCC CGCAGGGTG CCTACCCGCA  
 1251 GCCTGCATCG TTTGCCAAA CCATCCCGCA ATACGGCACC AGCGGCGAAA  
 1301 TCGGCGGCTA TCTGCCACC CGTTTCCGCG CCGCCGACAA CCTTTCGCTG  
 1351 ATACTCGGCG GCAGATACAG CCGTTACCGC ACCGGCAGCT ACGACAGCCG  
 1401 CACACAAGGC ATGACCTATG TGTCCGCCAA CCGTTTCACC CCCTACACAG  
 30 1451 GCATCGTGT CGACCTGACC GGCAACCTGT CGCTTTACGG CTGTACAGC  
 1501 AGCCTGTTG TCCCGCAATC GCAAAAAGAC GAACACGGCA GCTACCTGAA  
 1551 ACCCGTAACC GGCAACAATC TGGAAGCCGG CATCAAAGGC GAATGGCTTG  
 1601 AAGGCCGTCT GAACGCATCC GCCGCCGTGT ACCGCGCCCG TAAAAACAAC  
 35 1651 CTCGCCACCG CAGCAGGACG CGACCCGAGC GGCAACACCT ACTACCGCGC  
 1701 CGCCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GCGGCGCGCA  
 1751 TCACGCCCCA ATGGCAGATA CAGGCAGGTT ACAGCCAAG CAAAACCCGC  
 1801 GACCAAGACG GCAGCCGCTT GAACCCGAC AGCGTACCCG AACGCAGCTT  
 1851 CAAACTCTT ACTGCCTACC ACTTTGCCCC CGAAGCCCCC AGCGGCTGGA  
 40 1901 CCATCGGCGC AGGCGTGCGC TGGCAGAGCG AAACCCACAC CGACCTGCC  
 1951 ACGCTCCGCA TCCCCAACCC CGCCGCCAAA GCCCGCGCCG CCGACAACAG  
 2001 CCGCCAAAAA GCCTACGCCG TCGCCGACAT CATGGCGCGT TACCGCTTCA  
 2051 ATCCGCGCGC CGAACTGTG CTGAACGTGG ACAATCTGTT CAACAAACAC  
 2101 TACCGCACCC AGCCCGACCG CCACAGCTAC GCGCAGTGC GGACAGTGAA  
 2151 CGCGGCGTTT ACCTATCGGT TTAAATAA

45 This encodes a protein having amino acid sequence <SEQ ID 668>:

1 MTRFKYSLLE AALLPVYAQA DVSVDPPKP QESTELPTIT VTADRTASSN  
 51 DGYTVSGTHT PLGLPMTLRE IPQSVSVITS QOMRDQNIKA LDRALLQATG  
 101 TSRQIYGS DR AGYNLYFARG SRIANYQING IPVADALADT GNANTAAYER  
 50 151 VEVVRGVAGL LDGTGEPSAT VNLVRKRPR KPLFEVRAEA GNRKHFLGLA  
 201 DVSGLNAEG TLRGLVSTF GRGDSWRQRE RSRDAELYGI LEYDIAPQTR  
 251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWANSRHRAL  
 301 NLFAGIEHRE NQDWKLKAEY DYTRSFRFRQ YGVAGVLSID HNTAATDLIP  
 351 GYWHADPRTH SASVSLIGKY RLFGREHDLI AGINGYKYAS NKYGERSIIP  
 401 NAI PNAYEFS RTGAYQPAS FAQTIPOYGT RRQIGGYLAT RFRAADNLSL  
 55 451 ILGGYRSRYR TGSYDSRTQG MTYVSANRET PYTGIVFDLT GNLSLYGSYS  
 501 SLFVPQSQKD EHGSYLKPV T GNNLEAGIKG EWLEGRNLAS AAVYRARKNN  
 551 LATAAGRDP S GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKTR  
 601 DQDGSRLNPD SVPERSFKLF TAYHFAPEAP SGWTIGAGVR WQSEHTDPA  
 651 TLRIPNPAK ARAADNSRQK AYAVADIMAR YRFNPRAELS LNVNLFENKH  
 60 701 YRTQPDHRSY GALRTVNAAF TYREK\*

ORF23a and ORF23-1 show 99.2% identity in 725 aa overlap:

		10	20	30	40	50	60
orf23a.pep		MTRFKYSLLEAALLPVYAQA	DVSVDPPKPQESTELPTIT	VTADRTASSN	DGYTVSGTHT		
65	orf23-1	MTRFKYSLLEAALLPVYAQA	DVSVDPPKPQESTELPTIT	VTADRTASSN	DGYTVSGTHT		
		10	20	30	40	50	60

-382-

		70	80	90	100	110	120
	orf23a.pep	PLGLPMTLREIPQSVSVITSQOMRDQNIKALDRALLQATGTSRQIYGS DRAGYNL FARG					
5	orf23-1	PLGLPMTLREIPQSVSVITSQOMRDQNIKTLD RALLQATGTSRQIYGS DRAGYNL FARG					
		70	80	90	100	110	120
	orf23a.pep	130	140	150	160	170	180
10	orf23-1	SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEPSATVNLVRKRPTR					
		130	140	150	160	170	180
	orf23a.pep	190	200	210	220	230	240
15	orf23-1	KPLFEVRAEAGNRKHFGLGADVSGSLNAEGTLRGRLVSTFGRGDSWRQRERSRDAELYGI					
		190	200	210	220	230	240
	orf23a.pep	250	260	270	280	290	300
20	orf23-1	LEYDIAPQTRVHAGMDYQQAKETADAPLSYAVYDSQGYATAFGPKDNPATNWANSRHRAL					
		250	260	270	280	290	300
25	orf23a.pep	310	320	330	340	350	360
	orf23-1	NLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTH					
30		310	320	330	340	350	360
	orf23a.pep	370	380	390	400	410	420
	orf23-1	SASVSLIGKYRLFGREHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPAS					
35		370	380	390	400	410	420
	orf23a.pep	430	440	450	460	470	480
40	orf23-1	FAQTIPQYGTTRRQIGGYLATRFRAADNLSLILGGRYTRYRTGSYDSRTQGMTYVSANRFT					
		430	440	450	460	470	480
	orf23a.pep	490	500	510	520	530	540
45	orf23-1	PYTGIVFDLTGNLSLYGSYSSLFVPSQKDEHGSYLKPVTGNNLEAGIKGEWLEGRLNAS					
		490	500	510	520	530	540
50	orf23a.pep	550	560	570	580	590	600
	orf23-1	AAVYRARKNNLATAAGRDPSGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKTR					
		550	560	570	580	590	600
55	orf23a.pep	610	620	630	640	650	660
	orf23-1	DQDGSRLNPDSVPERSEFKLFTAYHFAPEAPSGWTIGAGVRWQSEHTDPATLRIPNPAK					
60		610	620	630	640	650	660
	orf23a.pep	670	680	690	700	710	720
	orf23-1	ARAADNSRQKAYAVADIMARYRFPRAELSLNVDNLFNKHYRTQPD RHSYGALRTVNAAF					
65		670	680	690	700	710	720
	orf23a.pep	TYRFXK					
70	orf23-1	TYRFXK					

Homology with a predicted ORF from *N.gonorrhoeae*

ORF23 shows 93.4% identity over a 211aa overlap with a predicted ORF (ORF23.ng) from *N.gonorrhoeae*:

5	orf23.pep	GYNYLFARGSR IANYQINGIPVADALADTGNANTAAYERVEVVVRGVAGLLD	51
	orf23ng	SAVDACRIPGYNYLFARGSR IANYQINGIPVADALADTGNANTAAYERVEVVVRGVAGLPD	60
10	orf23.pep	GTGEPSATVNLVRKRLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEXXLRGLVSTFGR	111
	orf23ng	GTGEPSATVNLVRKHPTKPLFEVRAEAGNRKHFGLDADVSGSLNAEGTLRGLVSTFGR	120
15	orf23.pep	GDSWRRRERSRXAELYGILEYDIAPQTRVHAXMDYQQAKEADAPLSYAVYDSQGYATAF	171
	orf23ng	GDSWRQLERSRDAELYGILEYDIAPQTRVHAGMDYQQAKEADAPLSYAVYDSQGYATAF	180
	orf23.pep	GPKDNPATNWANSHHRLNLFAGIEHFRNQDWKLKAEYDY	211
	orf23ng	GPKDNPATNWSNSRNRALNLFAGIEHFRNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHS	240

The ORF23ng nucleotide sequence <SEQ ID 669> is predicted to encode a protein comprising amino acid sequence <SEQ ID 670>:

	1	SAVDACRIPG	YNYLFARGSR	IANYQINGIP	VADALADTGN	ANTAAYERVE
	51	VVRGVAGLPD	GTGEPSATVN	LVRKHPTKRP	LFEVRAEAGN	RKHFGLDADV
	101	SGSLNAEGTL	RGRLVSTFGR	GDSWRQLERS	RDAELYGILE	YDIAPQTRVH
25	151	AGMDYQQAKE	TADAPLSYAV	YDSQGYATAF	GPKDNPATNW	SNSRNRALNL
	201	FAGIEHFRNQ	DWKLKAEYDY	TRSRFRQPYG	VAGVLSIDHS	TAATDLIPGY
	251	WHADPRTHSA	SMSLTGKYRL	FGREHDLIAG	INGYKYASNK	YGERSIIPNA
	301	IPNAYEFSRT	GAYPOPSFSA	QTIPQYDTRR	QIGGYLATRF	RAADNLSLIL
	351	GGYRSRYRAG	SYNSRTQGMT	YVSANRFTPY	TGIVFDLTGN	LSLYGSYSSL
30	401	FVPQLQKDEH	GSYLKPVGTN	NLEADIKGEW	LEGRLNASAA	VYRARKNNLA
	451	TAAGRDQSGN	TYYRAANQAK	THGWEIEVGG	RITPEWQIQG	GYSQSKPRDQ
	501	DGSRNPDSV	PERSFKLFTA	YHLAPEAPSG	RTIGAGVRRQ	GETHTDPAAL
	551	RIPNPAAKAR	AVANSRQKAY	AVADIMARYR	FNPRTELSLN	VDNLFNKHyr
	601	TQPDHRSYGA	LRTVNAAFTY	RFK*		

Further work revealed the complete nucleotide sequence <SEQ ID 671>:

35	1	ATGACACGCT	TCAAATACTC	CCTGCTTTTT	GCCGCCCTGC	TACCCGTGTA
	51	CGCGCAGGCC	GATGTTTCTG	TTTCAGACGA	CCCCAAACCG	CAGGAAAGCA
	101	CCGAATTGCC	GACCATCACC	GTTACCGCCG	ACCGCACCGC	GAGTTCCAAC
	151	GACGGCTACA	CCGTTTCCGG	CACGCACACC	CCGTTCCGGC	TGCCCATGAC
	201	CCTGCCGCAA	ATCCCGCAGA	GCGTCAGCGT	CATCACATCG	CAACAAATGC
40	251	GCGACCAAAA	CATCAAAACG	CTCGACCGCG	CCCTGTTGCA	GGCGACCGGC
	301	ACCAGCCGCC	AGATTTACGG	CTCCGACCGC	GCGGGCTACA	ACTACCTGTT
	351	CGCGCGCGGC	AGCCGCATCG	CCAATAACCA	AATCAACGGC	ATCCCCGTG
	401	CCGACGCGCT	GGCCGATACG	GGCAATGCCA	ACACCGCCGC	CTATGAGCGC
	451	GTAGAAGTCG	TGCGCGGCGT	GGCGGGGCTG	CCGACGGCA	CGGGCGAGCC
45	501	TTCTGCCACC	GTCAATCTGG	TACGCAAACA	CCCGACCGCG	AAGCCATTGT
	551	TTGAAGTCCG	CGCCGAAGCC	GGCAACCGCA	AACATTTCCG	GCTGGGCGCG
	601	GACGTATCGG	GCAGCCTGAA	CGCCGAAGGC	ACGCTGCGCG	GCCGCCTGGT
	651	TTCCACCTTC	GGACGCGGCG	ACTCGTGGCG	GCAGCTCGAA	CGCAGCCGCG
	701	ATGCCGAACT	CTACGGCATT	TTGGAATACG	ACATCGCACC	GCAAACCGCG
50	751	GTCCACGAG	GCATGGACTA	CCAGCAGGCG	AAAGAAACCG	CAGACGCGCC
	801	GCTCAGCTAC	GCCGTGTACG	ACAGCCAAGG	TTATGCCACC	GCCTTCGCCC
	851	CAAAAGACAA	CCCCGCCACA	AATTGGTCGA	ACAGCCGCAA	CCGTGCGCTC
	901	AACCTGTCG	CCGGCATAGA	ACACCGCTTC	AACCAAGACT	GGAAACTCAA
	951	AGCCGAATAC	GACTACACCC	GTAGCCGCTT	CCGCCAGCCC	TACGGTGTGG
55	1001	CAGGCGTACT	TTCATCGAC	CACAGCACTG	CCGCCACCGA	CCTGATTCCC
	1051	GGTTATTGGC	ACGCGgatcc	GCGCACCCAC	AGCGCCAGCA	TGTCATTGAC
	1101	CGGCAATATC	CgcctGTTCC	GCCGCGAGCA	CGATTTAATC	GCGGGTATCA
	1151	ACGGCTACAA	ATACGCCAGC	AACAAATACG	GCGAACGCAG	CATCATTTCCC
	1201	AACGCCATTC	CCAACGCCTA	CGAATTTTCC	CGCAGGGGCG	CCTATCCGCA
60	1251	GCCATCATCG	TTTGCCCAAA	CCATCCCGCA	ATACGACACC	AGGCGGCAAA
	1301	TCGGCGGCTA	TCTCGCCACC	CGTTTCCGCG	CCGCCGACAA	CCTTTCGCTG
	1351	ATACTCGGCG	GCAGATACAG	CCGCTACCGC	GCAGGCAGCT	ACAACAGCCG

-384-

1401 CACACAAGGC ATGACCTATG TGTCCGCCAA CCGTTTCACC CCCTACACAG  
 1451 GCATCGTGTT CGATCTGACC GGCAACCTGT CGCTTTACGG CTCGTACAGC  
 1501 AGCCTGTTCG TCCCGCAATT GCAAAAAGAC GAACACGGCA GCTACCTGAA  
 1551 ACCCGTAACC GGCAACAATC TGGAAGCCGA CATCAAAGGC GAATGGCTTG  
 1601 AAGGGCGTCT GAACGCATCC GCCGCCGTGT ACCGCGCCCG TAAAAACAAC  
 1651 CTCGCCACCG CAGCAGGACG CGACCAGAGC GGCAACACCT ACTATCGCGC  
 1701 CGCCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GGCGGCCGCA  
 1751 TCACGCCCGA ATGGCAGATA CAGGCAGGCT ACAGCCAAAG CAAACCCCGC  
 1801 GACCAAGACG GCAGCCGCCT GAACCCCGAC AGCGTAcCCG AACGCAGCTT  
 1851 CAAACTCTTC ACCGCTACC ACTTAGCCCC CGAAGCCCC AGCGGCCGGA  
 1901 CCATcggTGC GGGTGTGCGC CGCAGGGCG AAACCCACAC CGACCCAGCC  
 1951 GCGCTCCGCA TCCCAACCC CGCCGCCAAA GCCCGCGCG TCGCCAACAG  
 2001 CCGCCAGAAA GCCTACGCCG TCGCCGACAT CATGGCGCGT TACCGCTTCA  
 2051 ATCCGCGCAC CGAACTGTGCG CTGAACGTGG ACAACCTGTT CAACAAACAC  
 2101 TACCGCACCC AGCCCGACCG CCACAGCTAC GCGCAGTGC GGACAGTGAA  
 2151 CGCGGCGTTT ACCTATCGGT TTAAATAA

This corresponds to the amino acid sequence <SEQ ID 672; ORF23ng-1>:

1 MTRFKYSLLF AALLPVYAQA DVSVSDPKP QESTELPTIT VTADRTASSN  
 51 DGYTVSGTHT PFGLPMTLRE IPQSVSVITS QQMRDQNIKT LDRALLQATG  
 101 TSRQIYGS DR AGYNLFARG SRIANYQING IPVADALADT GNANTAAYER  
 151 VEVVRGVAGL PDGTGEPSAT VNLVRKHPT R KPLFEVRAEA GNRKHFGLGA  
 201 DVSGSLNAEG TLRGRLVSTF GRGDSWRQLE RSRDAELYGI LEYDIAPQTR  
 251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWSNSRNRAL  
 301 NLFAGIEHRF NQDWKLKAEY DYTRSFRQPYGVAGVLSIDHSTAATDLIP  
 351 GYWHADPRTH SASMSLTGKY RLFGRHDLI AGINGYKYAS NKYGERSIIP  
 401 NAIPNAYEFS RTGAYPQPSS FAQTIPQYDT RRQIGGYLAT RFRAADNLSL  
 451 ILGGYRSYR AGSYNSRTQG MTYVSANRET PYTGIVFDLT GNLSLYGSYS  
 501 SLFVPQLQKD EHGSYLKPV T GNNLEADIKG EWLEGRINAS AAVYRARKNN  
 551 LATAAGRDQS GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKPR  
 601 DQDGSRLNPD SVPERSFKLF TAYHLAPEAP SGRITIGAGVR RQGETHTDPA  
 651 ALRIPNPAK ARAVANSRQK AYAVADIMAR YRFNPRTELS LNVDNLFNKH  
 701 YRTQPDHRSY GALRTVNAAF TYRFFK\*

ORF23ng-1 and ORF23-1 show 95.9% identity in 725 aa overlap:

35	orf23-1.pep	10	20	30	40	50	60
		MTRFKYSLLF AALLPVYAQA DVSVSDPKP QESTELPTIT VTADRTASSN DGYTVSGTHT					
	orf23ng-1	MTRFKYSLLF AALLPVYAQA DVSVSDPKP QESTELPTIT VTADRTASSN DGYTVSGTHT					
40	orf23-1.pep	70	80	90	100	110	120
		PLGLEMTLREIPQSVSVITSQQMRDQNIKT LDRALLQATGTSRQIYGS DRAGYNLFARG					
	orf23ng-1	PFGLPMTLREIPQSVSVITSQQMRDQNIKT LDRALLQATGTSRQIYGS DRAGYNLFARG					
45	orf23-1.pep	130	140	150	160	170	180
		SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEPSATVNLVRKRLTR					
50	orf23ng-1	SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLPDGTGEPSATVNLVRKHPT R					
55	orf23-1.pep	190	200	210	220	230	240
		KPLFEVRAEAGNRKHFGLDADVSGSLNTEGTLRGRLVSTFGRGDSWRRRERSRDAELYGI					
	orf23ng-1	KPLFEVRAEAGNRKHFGLDADVSGSLNAEGTLRGRLVSTFGRGDSWRQLERSRDAELYGI					
60	orf23-1.pep	250	260	270	280	290	300
		LEYDIAPQTRVHAGMDYQQA KETADAPLSYAVYDSQGYATAFGPKDNPATNWSNSRNRAL					
	orf23ng-1	LEYDIAPQTRVHAGMDYQQA KETADAPLSYAVYDSQGYATAFGPKDNPATNWSNSRNRAL					
65	orf23-1.pep	310	320	330	340	350	360
		NLFAGIEHRFNQDWKLKAEYDYTRSFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTH					
	orf23ng-1	NLFAGIEHRFNQDWKLKAEYDYTRSFRQPYGVAGVLSIDHSTAATDLIPGYWHADPRTH					

-385-

		310	320	330	340	350	360
		370	380	390	400	410	420
5	orf23-1.pep	SASVSLIGKYRLFGREHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPAS					
	orf23ng-1	SASMSLTGKYRLFGREHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPSS					
		370	380	390	400	410	420
10	orf23-1.pep	430	440	450	460	470	480
	orf23ng-1	FAQTIPQYGTTRRQIGGYLATRFRAADNLSLILGGRYTRYRTGSYDSRTQGMITYVSANRFT					
		430	440	450	460	470	480
15	orf23-1.pep	490	500	510	520	530	540
	orf23ng-1	PYTGIVFDLTGNLSLYGSYSSLFVPSQKDEHGSYLKPVGTGNLEAGIKGEWLEGRNLAS					
		490	500	510	520	530	540
20	orf23-1.pep	550	560	570	580	590	600
	orf23ng-1	AAVYRARKNNLATAAGRDPSGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKTR					
25		550	560	570	580	590	600
	orf23-1.pep	610	620	630	640	650	660
	orf23ng-1	DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSGWTIGAGVRWQSEHTDTPALTRIPNPAK					
30		610	620	630	640	650	660
	orf23-1.pep	670	680	690	700	710	720
	orf23ng-1	ARAADNSRQKAYAVADIMARYRFNPRAEISLNVNLFNKHRYTQPDHRSYGALRTVNAAF					
35		670	680	690	700	710	720
40	orf23-1.pep	TYREFKX					
	orf23ng-1	TYREFKX					

In addition, ORF23ng-1 shows significant homology with an OMP from *E.coli*:

45	sp P16869 FHUE_ECOLI OUTER-MEMBRANE RECEPTOR FOR FE(III)-COPROGEN, FE(III)-FERRIOXAMINE B AND FE(III)-RHODOTRULIC ACID PRECURSOR >gi 1651542 gnl PID d1015403 (D90745) Outer membrane protein FhuE precursor [Escherichia coli]
50	>gi 1651545 gnl PID d1015405 (D90746) Outer membrane protein FhuE precursor [Escherichia coli] >gi 1787344 (AE000210) outer-membrane receptor for Fe(III)-coprogen, Fe(III)-ferrioxamine B and Fe(III)-rhodotrulic acid precursor [Escherichia coli] Length = 729 Score = 332 bits (843), Expect = 3e-90 Identities = 228/717 (31%), Positives = 350/717 (48%), Gaps = 60/717 (8%)
55	Query: 38 TITVTADRTASSN--DGYTVSGTHTPFGLPMTLREIPQSVSVITSQQMRDQNIKTLDRL 95 T+ V TA + + Y+V+ T + MT R+IPQSV++++ Q+M DQ ++TL + Sbjct: 43 TVIVEGSATAPDDGENDYSVTSTSTAGTKMQMTQRDIPQSVTIVSQQRMEDQQLQTLGEVM 102
60	Query: 96 LQATGTSRQIYGSDRAGYNYLFARGSRANYQINGIP-----VADALDTGNANTAA 147 G S+ SDRA Y ++RG +I NY ++GIP + DAL+D A Sbjct: 103 ENTLGISKSQADSDRALY---YSRGFQIDNYMVDGIPTYFESRWNLGDALSDM-----AL 154
65	Query: 148 YERVEVVRGVAGLPDGTGEP SATVNLRKHPTKPLF-EVRAEAGNRKHFGLGADVSGSL 206 +ERVEVVRG GL GTG PSA +N+VRKH T + +V AE G+ AD+ L Sbjct: 155 FERVEVVRGATGLMTGTGNPSAAINMVRKHATSREFKGDVSAEYGSWNKERYVADLQSPL 214
70	Query: 207 NAEGTLRGRLVSTFGRGDSWRQLERSRDAELYGILEYDIAPQTRVHAGMDYQQAKETADA 266 +G +R R+V + DSW S GI++ D+ T + AG +YQ+ + Sbjct: 215 TEDGKIRARIVGGYQNNDSWLDRYNSEKTFSGIVDADLGDLTTLTSLAGYEYQRIDVNSPT 274
	Query: 267 PLSYAVYDSQGYATAFGPKDNPATNWSNSNRNALNLFAGIEHRFNQDWKLKAEYDYTRSR 326

-386-

+++ G + ++ + A +W+ + +F ++ +F W+ ++  
 Sbjct: 275 WGGLPRWNTDGSSNSYDRARSTAPDWAYNDKEINKVFMTLKQQFADTWQATLNATHSEVE 334  
 Query: 327 F--RQPYGVAGVLSIDHSTAA--TDLIPGY-----WHADPRTHSA-SMSLTGKYRLF 374  
 F + Y A V D ++ PG+ W++ R A + G Y LFG  
 Sbjct: 335 FDSKMMYVDAYVNKADGMLVGYPYSNYGPGFDYVGGTGWNSGKRKVDALDLFADGSYELFG 394  
 Query: 375 REHDLIAGINGYKYASNKYGER--SIIPNAIPNAYEFSRTGAYPQPSSFAQTIPQYDTRR 432  
 R+H+L+ G Y +N+Y +I P+ I + Y F+ G +PQ Q++ Q DT  
 Sbjct: 395 RQHNLMEFG-GSYSKQNNRYFSSWANIFDEIGSFYNFN--GNFPQTDWSPQSLAQDDTTH 451  
 Query: 433 QIGGYLATRFRAADNLSLILGGRYSRYRAGSYNSRTQGPTY--VSANRFTPYTGIVFDXXX 491  
 Y ATR AD L LILG RY+ +R + +TY + N TPY G+VFD  
 Sbjct: 452 MKSLYAATRVTLADPLHLILGARYTNWRVDT-----LTYSMEKNHTTPYAGLVFDIND 504  
 Query: 492 XXXXXXXXXXXXFPQLQKDEHGSYLKPVGTGNNLEADIKGEWLEGRNLASAAVYRARKNNL 551  
 F PQ +D G YL P+TGNN E +K +W+ RL + A++R ++N+  
 Sbjct: 505 NWSTYASYTSIFQPQNDRSSGKYLAPITGNNYELGLKSDWMNSRLTTTLAIFRIEQDNV 564  
 Query: 552 ATAAGR---DQSGNTYYRAANQAKTHGWEIEVGGRTPEWQIQAGYSQSKPRDQDGSRLN 608  
 A + G +G T Y+A + + G E E+ G IT WQ+ G ++ D +G+ +N  
 Sbjct: 565 AQSTGTPIPGSNGETAYKAVDGTVSKGVEFELNGAITDNWQLTFGATRYIAEDNEGNAVN 624  
 Query: 609 PDSVPERSFKLFTAYHLAPEAPSGRTIGAGVRRQGETHTDPAALRIPNPAKARAVANSR 668  
 P ++P + K+ET+Y L P P T+G GV Q +TD P RA  
 Sbjct: 625 P-NLPRTTVKMFYSYRL-PVMPE-LTVGGGVNWQNRVYTDTV-----TPYGTFR--+E 672  
 Query: 669 QKAYAVADIMARYRFPNRTSLNVDNLFNKHYRTQPDRLH-SYGALRTVNAAFTYRF 724  
 Q +YA+ D+ RY+ L NV+NLF+K Y T + YG R + TY+F  
 Sbjct: 673 QGSYALVDLFTYQVTKNFSLQGNVNNLFDKTYDTNVEGSIVYGTNRNFSITGTYQF 729

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF23-1 (77.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 15A shows the results of affinity purification of the His-fusion protein, and Figure 15B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 15C) and for ELISA (positive result). These experiments confirm that ORF23-1 is a surface-exposed protein, and that it is a useful immunogen.

#### Example 80

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 673>:

1 ATGCGCACGG CAGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC  
 51 GGCAATGATG CCGGAAATGG TGTGCGCGGG CGTGTGCGCG GGAACGGCAA  
 101 TCATATCCAA GCCGACCGAA CAAACGGCGG TCATGGCTTC GAGTTTGTC  
 151 AGCGTCAGcA CGCTGCTTC GCGGcGgCa ATCATACCTT CGTCTTCGGA  
 201 AACGGGGATA AACGcGCCAC TCAAACCCCC GACCGCGCTG GAAGCCATCA  
 251 TGCCGCCTTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTTGTG  
 301 CCGTGCGTAC CGCAGACGCT CAAGCCATT TnTTCAAGAA TGCGTGCCAC  
 351 TnAGTCGCCG ACGGGG..

This corresponds to the amino acid sequence <SEQ ID 674; ORF24>:

1 MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISKPTE QTAVMASSLS  
 51 SVSTPASAAA IIPSSSETGI NAPLKPPTAL EAIMPPFFTA SFSNAKAADV  
 101 PCVPQTLKPI XSRMRATXSP TG..

Further work revealed the complete nucleotide sequence <SEQ ID 675>:

```

      1  ATGCGCACGG  CAGTGGTTTT  GCTGTTGATC  ATGCCGATGG  CGGCTTCGTC
    51  GGCAATGATG  CCGGAAATGG  TGTGCGCGGG  CGTGTGCGCG  GGAACGGCAA
    101  TCATATCCAA  GCCGACCGAA  CAAACGGCGG  TCATGGCTTC  GAGTTTGTCC
    151  AGCGTCAGCA  CGCCTGCTTC  GGCGGCGGCA  ATCATACCTT  CGTCTTCGGA
    201  AACGGGGATA  AACGCGCCAC  TCAAACCCCC  GACCGCGCTG  GAAGCCATCA
    251  TGCCGCCTTT  TTTCACGGCA  TCGTTCAGCA  ATGCCAAAGC  TGCTGTTGTG
    301  CCGTGCGTAC  CGCAGACGCT  CAAGCCCATT  TCTTCAAGAA  TGGTGCCAC
    351  TGAGTCGCCG  ACGGCGGGGG  TCGGCGCCAG  CGACAAGTCG  AGAATACCAA
    401  ACGGGATATT  CAGCATTTTT  GAGGCTTCGC  GGCCGATGAG  TTCGCCACG
    451  CGGGTAATTT  TGAAAGCAGT  TTTCTTCACT  ACTTCCGCAA  CTTGGGTCAA
    501  TGTCGTTGCA  TCTGAATTTT  CCAACGCGGC  TTTTACGACA  CCTGGGCCGG
    551  ATAGCCGAC  ATTGATAACG  GCATCCGCTT  CGCCCGAACC  ATGAAACGCG
    601  CCCGCCATAA  ACGGGTTGTC  TTCCACCGCG  TTGCAGAACA  CGACAATTTT
    651  AGCGCAGCCG  AAACCTTCGG  GCGTGATTTC  CGCCGTGCGT  TTGACGGTTT
    701  CGCCCGCCAG  CTTGACCGCA  TCCATATTGA  TACCGGCACG  CGTACTGCCG
    751  ATATTGATGG  AGCTGCACAC  AATATCGGTA  GTCTTCATCG  CTTCGGGAAT
    801  GGAGCGGATT  AACACCTCAT  CCGAAGGCGA  CATCCCTTTT  TGCACCAACG
    851  CGGAAAAACC  GCCGATAAAA  GACACACCGA  TGGCTTTGGC  AGCTTTATCC
    901  AAAGTTTGCG  CCACGCTGAC  GTAA
  
```

This corresponds to the amino acid sequence <SEQ ID 676; ORF24-1>:

```

      1  MRTAVVLLLI  MPMAASSAMM  PEMVCAGVSP  GTAIISKPTE  QTAVMASSLS
    51  SVTTPASAAA  IIPSSSETGI  NAPLKPPTAL  EAIMPPFFTA  SFSNAKAAVV
    101  PCVPQTLKPI  SSRMRATESP  TAGVGASDKS  RIPNGIFSIF  EASRPMSSPT
    151  RVILKAVFFT  TSATSVNVVA  SEFSNAAFTT  PGPDTPTLIT  ASASEP*NA
    201  PAINGLSSTA  LQNTTILAQP  KPSGVISAVR  LTVSPASLTA  SILIPARVLP
    251  ILMELHTISV  VFIASGMERI  NTSSEGDIPF  CTNAEKPIK  DTPMALAALS
    301  KVCATLT*
  
```

Computer analysis of this amino acid sequence gave the following results:

### 30 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF24 shows 96.4% identity over a 307 aa overlap with an ORF (ORF24a) from strain A of *N. meningitidis*:

```

      10      20      30      40      50      60
    35  orf24a.pep  MRTAVVLLLIMPMMAASSAMMP  PEMVCAGVSPGTAIISKPTEQTAVIASSLSNVSTPASAAA
      |||
    orf24  MRTAVVLLLIMPMMAASSAMMP  PEMVCAGVSPGTAIISKPTEQTAVMASSLSNVSTPASAAA
      |||
      10      20      30      40      50      60

      70      80      90     100     110     120
    40  orf24a.pep  IIPSSSXTGINAPLKPPTALEAIMPPFFFTASFSNAKAAVVPCVPQTLKPISSRMRATESP
      |||
    orf24  IIPSSSETGINAPLKPPTALEAIMPPFFFTASFSNAKAAVVPCVPQTLKPISSRMRATESP
      |||
      70      80      90     100     110     120

      130     140     150     160     170     180
    45  orf24a.pep  TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT
      |||
    orf24  TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT
      |||
      130     140     150     160     170     180

      190     200     210     220     230     240
    50  orf24a.pep  PGPDTPTLITASASPEPXNAPAIIXGLSSXALQNTTILAQPKPSSVISXVRLMVSPASLTA
      |||
    orf24  PGPDTPTLITASASPEPXNAPAIINGLSSTALQNTTILAQPKPSGVISAVRLTVSPASLTA
      |||
      190     200     210     220     230     240

      250     260     270     280     290     300
    55  orf24a.pep  SILIPARVLPILMELHTISVVFIASGMERXNTSSEGDIPFCTSAEKPIKDTPMALAALS
      |||
    orf24  SILIPARVLPILMELHTISVVFIASGMERINTSSEGDIPFCTNAEKPIKDTPMALAALS
      |||
      250     260     270     280     290     300
  
```



orf24a.pep KVCATLTX  
 |||||  
 orf24 KVCATLTX

5 The complete length ORF24a nucleotide sequence <SEQ ID 677> is:

1 ATGCGCACGG CAGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC  
 51 GGCAATGATG CCGGAAATGG TGTGCGCGGG TGTGTCGCCG GGAACGGCAA  
 101 TCATATCCAA NCCGACCGAA CAAACGGCGG TCATCGCTTC GAGTTTATCC  
 151 AACGTCAGCA CGCCTGCTTC GCGGCGGGCA ATCATACCTT CGTCTTCGGA  
 10 201 NACGGGGATA AACGCGCCAC TCAAACCGCC AACCGCGCTC GAAGCCATCA  
 251 TGCCGCCCTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTTGTG  
 301 CCGTGCCTAC CGCAGACGCT CAAACCCATT TCTTCAAGAA TGCGCGCCAC  
 351 CGAGTCGCGG ACGGCAGGGG TCGGTGCCAG CGACAAGTCG AGAATACCAA  
 401 ACGGGATATT CAGCATTTTT GAGGCTTCGC GGCCGATGAG TTCGCCCACG  
 15 451 CGGGTAATTT TGAAGGCGGT TTTCTTCACA ACTTCGGCAA CTTCGGTCAA  
 501 TGTGCTTGCA TCCGAATTTT CCAACGCGGC TTTTACGACA CCCGGGCCGG  
 551 ATACGCCGAC ATTAATCACA GCATCCGCTT CGCCTGAGCC GTGAAACCGG  
 601 CCGGCCATAN ACGGTTTGTC TTCCNCCGCG TTGCAGAACG CGACGATTTT  
 651 GGCGCAGCCG AAACCTTCTA GTGTGATTTC ANCCGTGCGT TTGATGGTTT  
 20 701 CGCCCCCAG TCTGACCGCG TCCATATTGA TACCGGCGCG CGTACTGCCG  
 751 ATATTGATGG AGCTGCACAC GATATCAGTA GTCTTCATCG CTTCGGGAAT  
 801 GGAACGGATN AACACCTCGT CAGAAGGCGA CATACCTTTT TGCACCAGCG  
 851 CGGAAAAGCC GCCAATAAAA GACACGCCGA TGGCTTTGGC AGCCTTATCC  
 901 AAAGTTTGCG CCACGCTGAC GTAA

25 This encodes a protein having amino acid sequence <SEQ ID 678>:

1 MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISXPTE QTAVIASSLS  
 51 NVSTPASAAA IIPSSSXTGI NAPLKPPTAL EAIMPPFFTA SFSNAKAAVV  
 101 PCVPQTLKPI SSRMRATESP TAGVGASDKS RIPNGIFSIF EASRPMSSPT  
 151 RVILKAVFFT TSATSVNVVA SEFSNAAFTT PGPDPTTLIT ASASPEP\*NA  
 201 PAIXGLSSXA LQNTTILAQP KPSSVISXVR LMVSPASLTA SILIPARVLP  
 251 ILMELHTISV VFIASGMERX NTSSEGDIPF CTSAEKPPIK DTPMALAALS  
 301 KVCATLT\*

It should be noted that this protein includes a stop codon at position 198.

ORF24a and ORF24-1 show 96.4% identity in 307 aa overlap:

35 orf24a.pep 10 20 30 40 50 60  
 MRTAVVLLLIMPMAASSAMMPEMVCAGVSPGTAIISXPTEQTAVIASSLSNVSTPASAAA  
 orf24-1 10 20 30 40 50 60  
 MRTAVVLLLIMPMAASSAMMPEMVCAGVSPGTAIISKPTQTAVMASSLSNVSTPASAAA  
 40 orf24a.pep 70 80 90 100 110 120  
 IIPSSSXTGINAPLKPPTALEAIMPPFFTASFSNAKAAVVPCVPQTLKPISSRMRATESP  
 orf24-1 70 80 90 100 110 120  
 IIPSSSETGINAPLKPPTALEAIMPPFFTASFSNAKAAVVPCVPQTLKPISSRMRATESP  
 45 orf24a.pep 130 140 150 160 170 180  
 TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT  
 orf24-1 130 140 150 160 170 180  
 TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT  
 50 orf24a.pep 190 200 210 220 230 240  
 PGPDPTTLITASASPEPXNAPAI XGLSSXALQNTTILAQPKPSSVISXVRLMVSPASLTA  
 orf24-1 190 200 210 220 230 240  
 PGPDPTTLITASASPEPXNAPAI NGLSSTALQNTTILAQPKPSGVISAVRLTVSPASLTA  
 55 orf24a.pep 250 260 270 280 290 300  
 SILIPARVLPILMELHTISVVFIASGMERXNTSSEGDIPFCTSAEKPPIKDTPMALAALS  
 orf24-1 250 260 270 280 290 300  
 SILIPARVLPILMELHTISVVFIASGMERINTSSEGDIPFCTNAEKPPIKDTPMALAALS

ORF24 shows 96.7% identity over a 121 aa overlap with a predicted ORF (ORF24ng) from *N.gonorrhoeae*:

10	orf24.pep	MRTAVVLLLLIMPAASSAMMPPEMVCAGVSPGTAIISKPTQETAVMASSLSVSTPASAAA	60
	orf24ng	MRTAVVLLLLIMPAASSAMMPPEMVCAGVSPGTAIMSKPTQETAVMASSLSVNTPASAAA	60
15	orf24.pep	IIPSSSETGINAPLKPPTALEAIMPPFFFTASFSNAKAADVPCVPQTLKPISRMRTXSP	120
	orf24ng	IIPSSSETGINAPLKPPTALEAIMPPFFFTASFSNAKAADVPCVPQTLKPISSRMRTESP	120
	orf24.pep	TG	122
		I:	
20	orf24ng	TAGVGASDKSRMPNGIFISIFEASRPMSSPTRVILKAVFFFTTSATSVRLTASEFSSAALT	180

	1	ATGCCGACCGG	CGGTGGTTTT	GCTGTTGATC	ATGCCGATGG	CGGCTTCGTC
	51	GGCGATGATG	CCGGAATGG	TGTGCGCGGG	CGTGTGCGCG	GGAACGGCAA
25	101	TCATGTCCAA	ACCAACGGAG	CAGACGGCGG	TCATGGCTTC	GAGTTTGTCC
	151	AGCGTCAACA	CGCTGTCCCT	GGCGGCGGGA	ATCATACCTT	CGTCTTCGGA
	201	AACGGGGATA	AACGCGCGC	TCAAACCGCC	GACCGCGCTG	GAAGCCATCA
	251	TGCCGCCCTT	TTTCACGGCA	TCGTTAGCA	ATGCCAAAGC	TGCTGTTGTG
	301	CGTGCGTAC	CGCAGACGCT	CAAGCCCATT	TCTTCAAGAA	TGCGCGCCAC
30	351	CGAGTCGCCG	ACGGCGGGGG	TCGTGCCAG	CGACAAATCG	AGAATGCCGA
	401	ACGGGATATT	CAGCATTTTT	GAGGCTTCGC	ACCCGATGAG	TTGCCCCACG
	451	CGGGTGATTT	TGAAAGCGGT	TTTCTTACAG	GACTTCGGCGA	CCTCGGTFCAG
	501	GCTGACCGCG	TCCGAATTTT	CCAGCGCGGC	TTTGACCACG	CCTGGACCGG
	551	ATACGCCGAC	ATTAATCACA	GCATCCGCTT	CGCCCGAGCC	GTGGAACGCA
35	601	CCCCCATAA	ACGGATTGTC	TTCCACCGCG	TTGCAGAACA	CGACGATTTT
	651	GGCGCAGCCG	AAACCTTCGG	GTGTGATTT	AGCCGTCGCT	TTGATGTGTT
	701	CGCCTGCCAG	CTTGACCGCA	TCCATATTGA	TACCGGCGACG	CGTGTGTCGG
	751	ATATTGATGG	AGCTGCACAC	GATATCGGTA	GTTTTTCATCG	CTTCGGGAAC
	801	GGAACGGATC	AACACCTCAT	CCGAAGGCGA	CATACCTTTT	TGCACGACGG
40	851	CGGAAAAAGC	GCCGATAAAG	GACACGCCGA	TGGCTTTGGC	TGCCTTGTCC
	901	AAAGTCTGCG	CCACGCTGAC	ATAA		

45

1	<u>MRTAVVLLLI</u>	<u>MPMAASSAMM</u>	PEMVCAGVSP	GTAIMSKPTE	QTAVMASSLS
51	SVNTPASAAA	IIPSSSETGI	NAPLKPPTAL	EAIMPPFFTA	SFSNAKAAVV
101	PCVPQTLKPI	SSRMRATESP	TAGVGASDKS	RMPNGIFISF	EASRPMSSPT
151	RVILKAVFFT	TSATSVRLTA	SEFSSAALT	PGPDPTTLTI	ASASPEEPWA
201	PAINGLSSTA	LQNTTILAQP	KPSGVISAVR	<u>LMVSPASLTA</u>	<u>SILIPARVLP</u>
251	<u>ILMELHTISV</u>	<u>VFIASGTERI</u>	NTSSEGDIPF	CTSAEKPPIK	DTPMALAALS
301	KVCATLT*				

50		10	20	30	40	50	60
	orf24-1.pep	MRTAVVLL	LIMPMAASSAM	PEMVCAGVSP	GTAIIISKPT	EQTAVMASS	LSVVSTPASAAA
	orf24ng						
		10	20	30	40	50	60
	orf24-1.pep	MRTAVVLL	LIMPMAASSAM	PEMVCAGVSP	GTAIMSKPT	EQTAVMASS	LSVVNTPASAAA
	orf24ng						
55		10	20	30	40	50	60
	orf24-1.pep	IIPSSSET	GINAPLKPPT	ALEAIMPPFF	TASFNSAKA	AAVVPCVP	QTLKPISSRM
	orf24ng						
		70	80	90	100	110	120
	orf24-1.pep	IIPSSSET	GINAPLKPPT	ALEAIMPPFF	TASFNSAKA	AAVVPCVP	QTLKPISSRM
	orf24ng						
60		70	80	90	100	110	120
	orf24-1.pep	IIPSSSET	GINAPLKPPT	ALEAIMPPFF	TASFNSAKA	AAVVPCVP	QTLKPISSRM
	orf24ng						

-390-

			130	140	150	160	170	180
	orf24-1.pep		TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTTSATSVNVVASEFSNAFFT					
			:				:	:
5	orf24ng		TAGVGASDKSRMPNGIFSIFEASRPMSSPTRVILKAVFFTTTSATSVRLTASEFSSAALT					
			130	140	150	160	170	180
			190	200	210	220	230	240
	orf24-1.pep		PGPDTPTLITASASPEPXNAPAINGLSSTALQNTTILAQPKPSGVISAVRLTVSPASLTA					
10	orf24ng		PGPDTPTLITASASPEPWNAPAINGLSSTALQNTTILAQPKPSGVISAVRLMVSPASLTA					
			190	200	210	220	230	240
			250	260	270	280	290	300
	orf24-1.pep		SILIPARVLPILMELHTISVVFIASGMERINTSSEGDIPFCTNAEKPPIKDTPMALAALS					
						:		
15	orf24ng		SILIPARVLPILMELHTISVVFIASGTERINTSSEGDIPFCTSAEKPPIKDTPMALAALS					
			250	260	270	280	290	300
20	orf24-1.pep	KVCATLT						
	orf24ng	KVCATLT						

Based on this analysis, including the presence of a putative leader sequence (first 18 aa – double-underlined) and putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 81

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 681>:

30 1 ..ACCGACGTGC AAAAAGAGTT GGTCCGCGAA CAACGCAAGT GGGCGCAGGA  
51 AAAAATCAGC AACTGCCGAC AAGCCGCCGC GCAGGCAGAC CGGCAGGAAT  
101 ACGCCGAATA CCTCAAGCTG CAATGCGACA CGCGGATGAC GCGCGAACGG  
151 ATACAGTATC TTCGCGGCTA TTCCATCGAT TAG

This corresponds to the amino acid sequence <SEQ ID 682; ORF25>:

35 1 ..TDVQKELVGE QRKWAQEKIS NCRQAAAQAD RQEYAEYLKL QCDTRMTRER  
51 IQYLRGYSID \*

Further work revealed the complete nucleotide sequence <SEQ ID 683>:

	1	ATGTATCGGA	AACATATTGC	GCTGCCGTTT	GCCCTGCTGC	TTGCCGCTTG
40	51	CGGCAGGGAA	GAACCGCCCA	AGGCATTGGA	ATGCGCCAAC	CCCGCCGTGT
	101	TGCAAGGCAT	ACGCGGCAAT	ATTACAGAAA	CGCTCACGCA	GGAAGCGCGT
	151	TCTTTCGCGC	GCGAAGACGG	CAGGCAGTTT	GTCGATGCCG	ACAAAAATTAT
	201	CGCGCGCCGC	TACGGTTTGG	CGTTTTCTTT	GGAACACGCT	TCGGAAACGC
	251	AGGAAGCGCG	GCGCACGTTT	TGTATCGCCG	ATTTGAACAT	TACCGTCCGC
45	301	TCTGAAACGC	TTGCCGATGC	CAAGGCAAAC	AGCCCCCTGT	TGTACGGGGA
	351	AACTGCTTTG	TCGGATATTG	TGCGGCAGAA	GACGGGCGGC	AATGTCGAGT
	401	TTAAAGACGG	CGTATTGACG	GCAGCCGTCC	GCTTCCTGCC	CGTCAAAGAC
	451	GGTCAGACGG	CATTGTGCGA	CAACACGGTC	GGTATGGCGC	CGCAAAACGT
	501	GTCTGCCGCG	CTGCTGCCTT	ACGGCGTGAA	GAGCATCTGT	ATGATAGACG
50	551	GCAAGGCGGT	GAAAAAAGAA	GACGCGGTCA	GGATTTTGAG	CGGAAAAGCC
	601	CGTGAAGAAG	AACCGTCCAA	ACCCACGCCC	GAAGACATTT	TGGAACACAA
	651	TGCCGCCGGC	GGCGATGCGG	GCGTACCCCA	AGCCGCAGAA	GGCGCGCCCG
	701	AACCGGAAAT	CCTGCATCCT	GACGACGGCG	AGCGTGCCCA	TACCGTTACC
	751	GTATCACGGG	GCGAAGTGGA	AGAGGCGCGC	GTACAAAAAC	AGCGGTGCGA
	801	ATCCGAAATT	ACCAAACCTT	GGGGAGGACT	CGATACCGAC	GTGCAAAAAG
55	851	AGTTGGTTCG	CGAACACGCG	AAGTGGGCGC	AGGAAAAAAT	CAGCAACTGC
	901	GACACAGCCG	CGCGCGAGGC	AGACCGGCAG	GAATACGCCG	AATACCTCAA
	951	GCTGCAATGC	GACACGCGGA	TGACGCGCGA	ACGGATACAG	TATCTTCGCG
	1001	GCTATTCCAT	CGATTAG			

This corresponds to the amino acid sequence <SEQ ID 684; ORF25-1>:

```

1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQGIRGN IQETLTQEAR
51 SFAREDGRQF VDADKIIAAA YGLAFSLEHA SETQEGGRTF CIADLNITVP
101 SETLADAKAN SPLLYGETAL SDIVRQKTGG NVEFKDGVLT AAVRFLPVKD
151 GQTAFVDNTV GMAAQTLASA LLPYGVKSIV MIDGKAVKKE DAVRILSGKA
201 REEPPSKPTP EDILEHNAAG GDAGVPQAAE GAPEPEILHP DDGERADTVT
251 VSRGEVEEAR VQNQRAESEI TKLWGGLDTD VQKELVGEQR KWAQEKISNC
301 RQAAAQADRQ EYAEYLKLCQ DTRMTREIRI YLRGYSID*

```

Computer analysis of this amino acid sequence gave the following results:

# 10 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF25 shows 98.3% identity over a 60aa overlap with an ORF (ORF25a) from strain A of *N. meningitidis*:

```

15 orf25.pep                      TDVQKELVGEQRKWAQEKISNCRQAAAQAD
    orf25a      VTVSRGEVEEARVQNQRAESEITKLWGGLDTDVQKELVGEQRKWAQEKISNCRQAAAQAD
                250      260      270      280      290      300

20 orf25.pep      RQEYAEYLKLCQDTRMTREIRIQYLRGYSIDX
    orf25a      RQEYAEYLKLCQDTRMTREIRIQYLRGYSIDX
                310      320      330

```

The complete length ORF25a nucleotide sequence <SEQ ID 685> is:

```

25 1 ATGTATCGGA AACTCATTGC GCTGCCGTTT GCCCTGCTGC TTGCCGCTTG
    51 CGGCAGGGAA GAACCGCCCA AGGCATTGGA ATGCGCCAAC CCCGCCGTGT
    101 TGCAANGCAT ACGCNGCAAT ATTCAGGAAA CGCTCACGCA GGAAGCGCGT
    151 TCCTTCGCGC GCGAAGACNG CANGCAGTTT GTCGATGCCG ACNAAATTAT
    201 CGCCGCCGCC TANGNTNNGN NGNTNTCTTT GGAACACGCT TCGGAAACGC
    251 AGGAAGGCGG GCGCACGTTT TGTNTCGCCG ATTTGAACAT TACCGTGCCG
    301 TCTGAAACGC TTGCCGATGC CAAGGCAAAAC AGCCCCCTGC TGTACGGGGA
    351 AACCGCTTTG TCGGATATTG TCGGCGAGAA GACGGGCGGC AATGTCGAGT
    401 TTAAGACCGG CGTATTGACG GCAGCCGTCC GCTTCCTACC CGTCAAAGAC
    451 GGTACAGANG CATTGTGCGA CAACACGGTC GGTATGGCGG CGCAAACGCT
    501 GTCTGCCCGC TTGCTGCCTT ACGGCGTGAA GAGCATCGTG ATGATAGACG
    551 GCAAGGCGGT AAAAAAAGAA GACGCGGTCA GGATTNTGAG CNGANAAGCC
    601 CGTGAANAAG AACCGTCCAA ANCCNNGCCC GAAGACATT TGGAAACATAA
    651 TGCCGCCGGA GGGGATGCAG ACGTACCCCA AGCCGGAGAA GACGCGCCCG
    701 AACCGGAAAT CCGTGCATCT GACGACGGCG AGCGTGCCGA TACCGTTACC
    751 GTATCACGGG GCGAAGTGA AGAGGCGCGN GTACAAAACC AGCGTGCCGA
    801 ATCCGAAATT ACCAAACTTT GGGGAGGACT CGATACCGAC GTGCAAAAAG
    851 AGTTGGTCGG CGAANAACGC AAGTGGGCGC AGGAAAAAAT CAGCAACTGC
    901 CGACAAGCCG CCGCGCAGGC AGACCGGCAG GAATACGCCG AATACCTCAA
    951 GCTGCAATGC GACACGCGGA TGACGCGCGA ACGGATACAG TATCTTCGCG
    1001 GCTATTCCAT CGATTAG

```

This encodes a protein having amino acid sequence <SEQ ID 686>:

```

1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQXIRXN IQETLTQEAR
51 SFAREDXXQF VDADKIIAAA XXXXXSLEHA SETQEGGRTF CXADLNITVP
101 SETLADAKAN SPLLYGETAL SDIVRQKTGG NVEFKDGVLT AAVRFLPVKD
151 GQXAFVDNTV GMAAQTLASA LLPYGVKSIV MIDGKAVKKE DAVRIXSXXA
201 REXEPSKXXP EDILEHNAAG GDADVPQAGE DAPEPEILHP DDGERADTVT
251 VSRGEVEEAR VQNQRAESEI TKLWGGLDTD VQKELVGEQR KWAQEKISNC
301 RQAAAQADRQ EYAEYLKLCQ DTRMTREIRI YLRGYSID*

```

ORF25a and ORF25-1 show 93.5% identity in 338 aa overlap:

```

55 orf25a.pep      MYRKLIALPFALLLAACGRE EPPKALECAN PAVLQXIRXNIQETLTQEARSFAREDXXQF
    orf25-1      MYRKLIALPFALLLAACGRE EPPKALECAN PAVLQGIRGNIQETLTQEARSFAREDGRQF

```

-392-

		10	20	30	40	50	60
		70	80	90	100	110	120
5	orf25a.pep	VDADXI	IAAAXXXXX	SLEHASETQEGGRTFCXADLNITVPSETLADAKANSPLYGETAL			
	orf25-1	VDADKI	IAAAYGLAFSLEHASETQEGGRTFCIADLNITVPSETLADAKANSPLYGETAL				
		70	80	90	100	110	120
10	orf25a.pep	SDIVRQKTGGNVEFKDGVLTAAVRFLPVKDGQXAFVDNTVGMAAQTL	SAALLPYGVKSIV				
	orf25-1	SDIVRQKTGGNVEFKDGVLTAAVRFLPVKDGQTAFVDNTVGMAAQTL	SAALLPYGVKSIV				
		130	140	150	160	170	180
15	orf25a.pep	MIDGKAVKKE	DAVRIXSXXAREXEPSKXXPEDILEHNAAGGDADVPQAGEDAPEPEILHP				
	orf25-1	MIDGKAVKKE	DAVRILSGKAREEESP	KPTPEDILEHNAAGGDAGVPPQAAEGAPEPEILHP			
		190	200	210	220	230	240
20	orf25a.pep	DDGERADTVTVSRGEVEEARVQNQRAESEITKLWGGLD	TDVQKELVGEQRKWAQEKISNC				
	orf25-1	DDGERADTVTVSRGEVEEARVQNQRAESEITKLWGGLD	TDVQKELVGEQRKWAQEKISNC				
		250	260	270	280	290	300
25	orf25a.pep	RQAAAQADRQ	EYAEYLKLC	QCDTRMTRE	RIQYLRGYSID		
	orf25-1	RQAAAQADRQ	EYAEYLKLC	QCDTRMTRE	RIQYLRGYSID		
		310	320	330	339		
30	orf25a.pep	RQAAAQADRQ	EYAEYLKLC	QCDTRMTRE	RIQYLRGYSID		
	orf25-1	RQAAAQADRQ	EYAEYLKLC	QCDTRMTRE	RIQYLRGYSID		
		310	320	330			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF25 shows 100% identity over a 60aa overlap with a predicted ORF (ORF25ng) from

35 *N.gonorrhoeae*:

	orf25.pep	TDVQKELVGEQRKWAQEKISNCRQAAAQAD	30
	orf25ng	VTVSERGEVEEARVQNQRAESEITKLWGGLD	TDVQKELVGEQRKWAQEKISNCRQAAAQAD 308
40	orf25.pep	RQEYAEYLKLCQDTRMTRE	RIQYLRGYSID 60
	orf25ng	RQEYAEYLKLCQDTRMTRE	RIQYLRGYSID 338

The complete length ORF25ng nucleotide sequence <SEQ ID 687> is:

45	1	ATGTATCGGA	AACTCATTGC	GCTGCCGTTT	GCCCTGCTGC	TTGCAGCGTG
	51	CGGCAGGGAA	GAACCGCCCA	AGGCGTTGGA	ATGCGCCAAC	CCGCGCGTGT
	101	TGCAGGACAT	ACGCGGCAGT	ATTGAGGAAA	CGCTCAGCA	GGAAGCGCGT
	151	TCTTTTCGCG	GCGAAGACGG	CAGGCAGTTT	GTCGATGCCG	ACAAAATTAT
	201	CGCCGCCGCC	TACGGTTTGG	CGTTTTCTTT	GGAACACGCT	TCGGAAACGC
	251	AGGAAGGCGG	GCGCACGTTT	TGTATCGCCG	ATTTGAACAT	TACCGTGCCG
50	301	TCTGAAACGC	TTGCCGATGC	CGAGGCAAC	AGCCCCCTGC	TGTATGGGGA
	351	AACGTCTTTG	GCAGACATCG	TGCAGCAGAA	GACGGGCGGC	AATGTCGAGT
	401	TTAAAGACGG	CGTATTGACG	GCAGCCGTC	GCTTCCTGCC	CGCCAAAGAC
	451	GCTCGGACGG	CATTATCGA	CAACACGGTC	GGTATGGCGA	CGCAAACGCT
	501	GTCTGCCGCG	TTGCTGCCTT	ACGGCGTGAA	GAGCATCGTG	ATGATAGACG
55	551	GCAAGGCGGT	GACAAAAGAA	GACGCGGTCA	GGGTTTGTAG	CGGCAAAGCC
	601	CGTGAAGAAG	AACCGTCCAA	ACCCACCCCG	GAAGACATTT	TGGAACACAA
	651	TGCCGCCGCG	GGCGATGCGG	GCGTACCCCA	AGCCGCAGAA	GGCGCACCCG
	701	AACCCGAAAT	CCTGCATCCC	GACGACGTCG	AGCGTGCCGA	TACCGTTACC
	751	GTATCACGGG	GCGAAGTGGG	AGAGGCGCGC	GTACAAAACC	AACGTGCGGA
60	801	ATCCGAAATT	ACCAAACTTT	GGGGAGGACT	CGATACCGAC	GTGCAAAAAG
	851	AGTTGGTCCG	CGAACAGCGC	AAGTGGGCGC	AGGAAAAAAT	CAGcaactgc
	901	cgACAAGCCG	CCGCGCAGGC	AGACCGGCAG	GAATACGCCG	AATACCTCAA
	951	GCTCCAATGC	GACACGCGGA	TGACGCGCGA	ACggaTACAG	TATCTTCGCG
	1001	GCTATTCCAT	CGATTAG			

65 This encodes a protein having amino acid sequence <SEQ ID 688>:

-393-

1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQDIRGS IOETILTQEAR  
 51 SFAREDGRQF VDADKIIAAA YGLAFSLEHA SETQEGGRTF CIADLNITVP  
 101 SETLADAEAN SPILYGETSL ADIVQKQTGG NVEFKDGVLT AAVRFLPAKD  
 151 ARTAFIDNTV GMATQTLAAL LPYGVKSIV MIDGKAVTKE DAVRVLGSKA  
 201 REEPPSKPTP EDILEHNAAG GDAGVPQAAE GAPEPEILHP DDVERADTVT  
 251 VSRGEVEEAR VQNQRAESEI TKLWGGLDTD VQKELVGEQR KWAQEKISNC  
 301 RQAAQADRQ EYAEYLKLCQ DTRMTRERIQ YLRGYSID\*

ORF25ng and ORF25-1 show 95.9% identity in 338 aa overlap:

10	orf25-1.pep	10	20	30	40	50	60
	orf25ng	10	20	30	40	50	60
15	orf25-1.pep	70	80	90	100	110	120
	orf25ng	70	80	90	100	110	120
20	orf25-1.pep	130	140	150	160	170	180
	orf25ng	130	140	150	160	170	180
25	orf25-1.pep	190	200	210	220	230	240
	orf25ng	190	200	210	220	230	240
30	orf25-1.pep	250	260	270	280	290	300
	orf25ng	250	260	270	280	290	300
35	orf25-1.pep	310	320	330	339		
	orf25ng	310	320	330	339		

45 Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF25-1 (37kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 16A shows the results of affinity purification of the GST-fusion protein, and Figure 16B shows the results of expression of the His-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 16C), ELISA (positive result), and FACS analysis (Figure 16D). These experiments confirm that ORF25-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 16E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF25-1.

### Example 82

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 689>

```

5      1  ATGCAGCTGA TCGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
      51  TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
     101  GCATCGGTAT TCTGGwysGC GTTGCCTTTT TGGTCGGCGG CAACCCCGTC
     151  GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTGGGTCAGA
     201  CGsyGATTGG TCGCTGGGCA AACCAAAAT CTGGGTTTTC CkGATACTTT
     251  TGGGTATTTT TACTTCCCTG CTGACCTACT CCGGCAGCAA T.....

10      851  .....AC TTGCTGGTA
      901  TTCGGCGGCA CTTGCGGCGT CTTTGCCGTC GTTCTCTGCA CGCTCGGCAC
     951  GATTAAACC GCCACTATC CCAAAGCCGT TTGGCAGGGT GCGAAATCTA
    1001  TGTTCCGGCG AATCGCCATT TTAATCCTCG CTTGGCTCAT CAGTACGGTT
    1051  GTCGGCGAAA TGCACACCGG CGATTACCTC TCCACACTGG TTGCGGGCAA
    1101  CATCCATCCC GGCTTCCTGC CCGTCATCCT CTTCTGCTC GCCAGCGTGA
    1151  TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCCG
    1201  ATTGCCGCGC CCATGGCGGT CAAAGTCGAA CCCGCGCTGA TTATCCCGTG
    1251  TATGTCCGCA GTAATGGCGG GGGCGGTATG CCGCGACCAC TGCTCGCCCA
    1301  TTTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC
    1351  GACCACGTTA CCTCGCAACT GCCTTACGCC TTAACCGTTG CCGCGCCCGC
    1401  CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGCT
    1451  TTGGCACGAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAAGAT
    1501  AAAAAA..
  
```

25 This corresponds to the amino acid sequence <SEQ ID 690; ORF26>:

```

      1  MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILXX VAFLVGGNPV
     51  DGLTHLKDMV VGLAWSDXDW SLGKPKILVF XILLGIFTSI LTYSGSN...

      251  .....TSLV
     301  FGGTCGVFAV VLCTIGTIKT ADYPKAVWQG AKSMFGAIAI LILAWLISTV
     351  VGEMHTGDYL STLVAAGNIHP GFLPVILFLL ASVMAFATGT SWGTFGIMLP
     401  IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARCNIH
     451  DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLFGTTGIV LAVLIFLLKD
     501  KK..
  
```

35 Further work revealed the complete nucleotide sequence <SEQ ID 691>:

```

      1  ATGCAGCTGA TCGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
     51  TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
    101  GCATCGGTAT TCTGGTCCGC GTTGCCTTTT TGGTCGGCGG CAACCCCGTC
    151  GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTGGGTCAGA
    201  CGGCGATTGG TCGCTGGGCA AACCAAAAT CTGGGTTTTC CTGATACTTT
    251  TGGGTATTTT TACTTCCCTG CTGACCTACT CCGGCAGCAA TCAGGCGTTT
    301  GCCGACTGGG CAAAACGGCA CATTAAAAAC CCGCGCGGCG CGAAATGCT
    351  GACCGCTGCG CTCGTGTTTC TAACCTTTAT CGACGACTAT TTCCACAGTC
    401  TCGCCGTCGG TCGGATTGCC CGCCCGTTA CCGACAAGTT TAAAGTTTCC
    451  CGCACCAAAC TCGCTACAT CCTCGACTCC ACTGCCGCTC CTATGTGCGT
    501  GCTGATGCCC GTTCAAGCT GGGGCGCGTC GATTATCGCC ACGCTTGCCG
    551  GACTGCTCGT TACCTACAAA ATCACCGAAT ACACGCCGAT GGGGACGTTT
    601  GTCGCCATGA GCCTGATGAA CTATTACGCA CTGTTTGCCC TGATTATGGT
    651  GTTCGTCGTC GCATGGTTTT CCTTCGACAT CGGCTCGATG GCACGTTTCG
    701  AACCAAGCCGC GTTGAACGAA GCCACGATG AAACCTGCCG TTCAGACGCT
    751  ACCAAAGGTC GTGTTTACGC ACTGATTATT CCCGTTTTGG CCTTAATCGC
    801  CTCAACGGTT TCCGCCATGA TCTACACCGG CCGCGAGGCA AGCGAAACCT
    851  TCAGCATTTT GGGGGCATTT GAAAACACGG ACGTAAACAC TTCGCTGGTA
    901  TTCGGCGGCA CTTGCGGCGT CCTTCCGCTC GTTCTCTGCA CGCTCGGCAC
    951  GATTAAACC GCCACTATC CCAAAGCCGT TTGGCAGGGT GCGAAATCTA
   1001  TGTTCCGGCG AATCGCCATT TTAATCCTCG CTTGGCTCAT CAGTACGGTT
   1051  GTCGGCGAAA TGCACACCGG CGATTACCTC TCCACACTGG TTGCGGGCAA
   1101  CATCCATCCC GGCTTCCTGC CCGTCATCCT CTTCTGCTC GCCAGCGTGA
   1151  TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCCG
   1201  ATTGCCGCGC CCATGGCGGT CAAAGTCGAA CCCGCGCTGA TTATCCCGTG
   1251  TATGTCCGCA GTAATGGCGG GGGCGGTATG CCGCGACCAC TGCTCGCCCA
   1301  TTTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC
  
```

-395-

1351 GACCACGTTA CCTCGCAACT GCCTTACGCC TTAACCGTTG CCGCCGCCGC  
 1401 CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTGGGGCT  
 1451 TTGGCAGCAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAAGAT  
 1501 AAAAAACGCG CCAACGCCTG A

5 This corresponds to the amino acid sequence <SEQ ID 692; ORF26-1>:

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFLVGGNPV  
 51 DGLTHLKDMV VGLAWSGDGW SLGKPKILVF LILLGIFTSL LTYSGSNQAF  
 101 ADWAKRHIKN RRGAKMLTAC LVEVTFIDY FHSLAVGAIA RPVTDKFKVS  
 151 RTKLAYILDS TAAPMCVLMF VSSWGASIIA TLAGLLVYTK ITEYTPMGTF  
 10 201 VAMSLMNYIA LEALIMVFVV AWFSFDIGSM ARFEQAALNE AHDETAVSDA  
 251 TKGRVYALII PVLALIASTV SAMIYTGAQA SETFSILGAF ENTDVNTSLV  
 301 FGGTCGVLA VCLTLGTIKT ADYPKAVWQG AKSMFGAIAI LILAWLISTV  
 351 VGENHTGDYL STLVAAGNIHP GFLEPVILFL ASVMAFATGT SWGTFGIMLP  
 401 IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDITIL SSTGARCNIH  
 15 451 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGTTGIV LAVLIFLLKD  
 501 KKRNA\*

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical transmembrane protein HI1586 of *H. influenzae* (accession number P44263)

ORF26 and HI1586 show 53% and 49% amino acid identity in 97 and 221 aa overlap at the

20 N-terminus and C-terminus, respectively:

Orf26 1 MQLIDYSHSFFSVVPPFLALALAVITRRVXXXXXXXXXXVAFLVGGNPVDGLTHLKDMV 60  
 M+LID+S S +S+VP LA+ LA+ TRRV L +L V  
 HI1586 14 MELIDFSSSVSIVPALLAILAIATRRLVSLSAGIIIGSLMLSDWQIGSAFNILVKNV 73  
 Orf26 61 VGLAWSDXDWSLKGPKILVFXILLGIFTSLTYSGSN 97  
 V L ++D + + I++F +LLG+ T+LLT SGSN  
 HI1586 74 VSLVYADGEIN-SNMNIVLFLLLGLVLTALLTVSGSN 109  
 //  
 Orf26 86 IFTSLTYSGS--NTSLVFGGTCGVFAVVLCTL--GTIKTADYPKAVWQGA KSMFGXXXX 141  
 +F+ L T+ + TSLV GG C + L + + +Y ++ G KSM G  
 HI1586 299 VFSVLGTFENTVVGTSVLVGGFCSSIIISTLLIILDRQVSVEPYVRSWIVGIKSMGSAIAI 358  
 Orf26 142 XXXXXXSTVVGEMHTGDYLSTLVAGNIHPGFLPVILFLLASVMAFATGT SWGTFGIMLP 201  
 + +VG+M TG YLS+LV+GNI FLPVILF+L + MAF+TGTSWGTFGIMLP  
 HI1586 359 LFFAWTINKIVGDMQTGKYLSSLVSGNIPMQFLPVILFVLGAAMAFSTGT SWGTFGIMLP 418  
 Orf26 202 IAAAMAVKVEPALIIPCMSAVMAGAVCGDH CSPISDITILSSTGARCNIHDHVTSQXXXX 261  
 IAAAMA P L++PC+SAVMAGAVCGDH CSP+SDTILSSTGA+CNHIDHVT+Q  
 HI1586 419 IAAAMAANAPELLPCLSAVMAGAVCGDH CSPVSDTILSSTGAKCNHIDHVT+Q LPYA 478  
 Orf26 262 XXXXXXXXXXXXXXXXSALLGFGTTGIVLAVLIFLLKDK 302  
 S L GF T + L V+IF +K +  
 45 HI1586 479 ATVATATSIGYIVGFTYSGLAGFAATAVSLIVIIFAVKKR 519

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF26 shows 58.2% identity over a 502aa overlap with an ORF (ORF26a) from strain A of *N.*

*meningitidis*:

50 orf26.pep 10 20 30 40 50 60  
 MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILXXVAFLVGGNPVDGLTHLKDMV  
 orf26a MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFVGGNPVDGLTHLKDMV  
 55 orf26.pep 70 80 90 99  
 VGLAWSDXDWSLKGPKILVFXILLGIFTSLTYSGSNXX-----  
 orf26a VGLAWSGDWSLKGPKILVFXILLGIFTSLTYSGSNQAFADWAKRHIKNRRGAKMLTAC  
 60 70 80 90 100 110 120



-396-

```

5  orf26.pep -----
   orf26a  LVEVTFIDDFHSLAVGAXARPVTDKFKVSRKLAYILDSTAAPMCVLMPVSSWGASIIA
           130      140      150      160      170      180

10  orf26.pep -----
   orf26a  TLAGLLVTTYKITEYTPMGTFFVAMSLMNYIALFALIMVFVVAWFSDIGSMARFEQAALNE
           190      200      210      220      230      240

15  orf26.pep -----
   orf26a  AHDETAUSDGSGWRVYALIIIPVLALIASTVSAMIYTTGAQASETFSILGAFENTDVNTSLV
           250      260      270      280      290      300
                                100      110
                                -----TSLV
                                ||||

20  orf26.pep 120      130      140      150      160      170
   orf26a  FGGTCGVFAVVLCTLGTIKTADYPKAVWQGA KSMFGAIAILILAWLISTVVGEMHTGDYL
           |||||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   orf26a  FGGTCGVLA VVLCTLGTIKIADYPKAVWQGA KSMFGAIAILILAWLISTVVGEMHTGDYL
           310      320      330      340      350      360

25  orf26.pep 180      190      200      210      220      230
   orf26a  STLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLP IAAAMAVKVEPALIIPCMSA
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   orf26a  STLVAGNIHPGFLXVILFLLASVMAFATGTSWGTFGIMLP IAAAMAVKVDPSLIIPCMSA
           370      380      390      400      410      420

30  orf26.pep 240      250      260      270      280      290
   orf26a  VMAGAVCGDHCSPISDTTILSSTGARNHIDHVTSQLPYALTVA AAAASGYLALGLTKSA
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   orf26a  VMAGAVCGDHCSPISDTTILSSTGARNHIDHVTSQLPYALTVA AAAASGYLALGLTKSA
           430      440      450      460      470      480

35  orf26.pep 300      310
   orf26a  LLGFGTTGIVLAVLIFLLKDKK
           |||||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
   orf26a  LLGFGXTGIVLAVLIFLLKDKK RANAX
           490      500

```

The complete length ORF26a nucleotide sequence <SEQ ID 693> is:

```

45  1  ATGCAGCTGA TCGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
    51  TTTGGCACTG GCACCTGCCG TCATTACCCG CCGCGTACTG CTGCTTTTAG
   101  GCATCGGTAT TCTGGTCGGC GTTGCCCTTT TGGTGCGCGG CAACCCGTC
   151  GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTTGGTCAGA
   201  CGGCGATTGG TCCTGGGCA AACC AAAANT CTGGT TTTCT CTGATACTTT
   251  TGGGTATTTT TACTTCCCTG CTGACCTACT CCGGCAGCAA TCAGGCGTTT
   301  GCCGACTGGG CAAAACGGCA CATTAAAAAC CGGCGCGGCG CGAAATGCT
   351  GACCGCCTGC CTCGTGTTCT TAACCTTTAT CGACGACTAT TTCCACAGTC
   401  TCGCCGTCGG TCGCNTTGCC CGCCCGTTA CCGACAAGTT TAAAGTTTCC
   451  CGCGCCAAAC TCGCCTACAT CCTCGACTCC ACTGCCGCGC CTATGTGCGT
   501  GCTGATGCCC GTTCAAGCT GGGGCGGCTC GATTATCGCC ACGCTTGCCG
   55  551  GACTGCTCGT TACCTACAAA ATCACC GAAT ACACGCCGAT GGGGACGTTT
   601  GTCGCCATGA GCCTGATGAA CTATTACGCA CTGTTTGCCC TGATTATGGT
   651  GTTCGTCGTC GCATGGTTCT CCTTCGACAT CGGCTCGATG GCACGTTTCG
   701  AACAAGCCGC GTTGAACGAA GCCCAGCATG AAAC TGCCGT TTCAGACGGC
   751  AGCTGGGCA GGGTTTACGC ATTGATTATT CCGTTTTTGG CCTTAATCGC
   801  CTCAACGGTT TCCGCCATGA TCTACACCGG TGCACAGGCA AGCGAAACCT
   851  TCAGCATTTT GGTGTCATT GAAAATACGG ACGTGAACAC TTCGCTGGTA
   901  TTCGGCGGCA CTGCGGCGT GCTTGCCGTC GTCTCTGCA CGCTCGGCAC
   951  GATTAAATC GCCGATTAT CCAAAGCCGT TTGGCAGGCT GCGAAATCCA
  1001  TGTTCGCGCG AATCGCCATT TTAATCCTTG CTTGGCTCAT CAGTACGGTT
  1051  GTCGGCGAAA TGCACACAGG CGACTACCTC TCCACGCTGG TTGCGGGCAA
  1101  CATCCATCCC GGCTTCCTGN CCGTCATCCT TTCTCTGCTC GCCAGCGTGA
  1151  TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTGCGCAT CATGCTGCCG
  1201  ATTGCCCGCG CCATGGCGGT CAAAGTCGAT CCCTCACTGA TTATCCCGTG
  1251  TATGTCCGCC GTGATGGCGG GGGCGGTATG CGGCGACCAC TGCTCGCCCA
  1301  TTTCCGACAC GACCATCCTG TCGTCCACCG GCGCGGCTG CAACCACATC

```

-397-

1351 GACCACGTTA CNTCGCAACT GCCTTACGCC TTAACCGTTG CCGCCGCCGC  
 1401 CGCATCGGGN TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGTT  
 1451 TTGGCANGAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAGAT  
 1501 AAAAAACGCG CCAACGCCTG A

5 This encodes a protein having amino acid sequence <SEQ ID 694>:

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFVLGGNPV  
 51 DGLTHLKDMV VGLAWSGDGW SLGKPKXLVF LILLGIFTSL LTYSGSNQAF  
 101 ADWAKRHIKN RRGAKMLTAC LVFVTFIDY FHS LAVGAXA RPVTDKFKVS  
 151 RAKLAYILDS TAAPMCVLMF VSSWGASIIA TLAGLLVITYK ITEYTPMGTF  
 10 VAMSLMNYIA LFALIMVFVV AWFSFDIGSM ARFEQAALNE AHDETAVSDG  
 201 SWGRVYALII PVLALIASTV SAMIYTGAQA SETFSILGAF ENTDVNTSLV  
 251 FGGTCGVLAV VLCTLTGIKI ADYPKAVWQG AKSMFGAIAI LILAWLISTV  
 301 VGEMHTGDYL STL VAGNIHP GFLXVILFLL ASVMAFATGT SWGTFGIMLP  
 351 TAAAMAVKVD PSLIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARCNIH  
 401 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLFGXGTGIV LAVLIFLLKD  
 15 451  
 501 KKRANA\*

ORF26a and ORF26-1 show 97.8% identity in 506 aa overlap:

		10	20	30	40	50	60
20	orf26a.pep	MQLIDYSHSF	FSVVPPFLAL	ALAVITRRVL	LSLGIGILVG	VAFVLGGNPV	DGLTHLKDMV
	orf26-1	MQLIDYSHSF	FSVVPPFLAL	ALAVITRRVL	LSLGIGILVG	VAFVLGGNPV	DGLTHLKDMV
		10	20	30	40	50	60
25	orf26a.pep	VGLAWSGDGW	SLGKPKXLVF	LILLGIFTSL	LTYSGSNQAF	ADWAKRHIKN	RRGAKMLTAC
	orf26-1	VGLAWSGDGW	SLGKPKXLVF	LILLGIFTSL	LTYSGSNQAF	ADWAKRHIKN	RRGAKMLTAC
		70	80	90	100	110	120
30	orf26a.pep	LVFVTFIDY	FHS LAVGAXA	RPVTDKFKVS	RAKLAYILDS	TAAPMCVLMF	VSSWGASIIA
	orf26-1	LVFVTFIDY	FHS LAVGAXA	RPVTDKFKVS	RAKLAYILDS	TAAPMCVLMF	VSSWGASIIA
		130	140	150	160	170	180
35	orf26a.pep	TLAGLLVITYK	ITEYTPMGTF	VAMSLMNYIA	LFALIMVFVV	AWFSFDIGSM	ARFEQAALNE
	orf26-1	TLAGLLVITYK	ITEYTPMGTF	VAMSLMNYIA	LFALIMVFVV	AWFSFDIGSM	ARFEQAALNE
		190	200	210	220	230	240
40	orf26a.pep	AHDETAVSDG	SWGRVYALII	PVLALIASTV	SAMIYTGAQA	SETFSILGAF	ENTDVNTSLV
	orf26-1	AHDETAVSDG	SWGRVYALII	PVLALIASTV	SAMIYTGAQA	SETFSILGAF	ENTDVNTSLV
		250	260	270	280	290	300
45	orf26a.pep	FGGTCGVLAV	VLCTLTGIKI	ADYPKAVWQG	AKSMFGAIAI	LILAWLISTV	VGEMHTGDYL
	orf26-1	FGGTCGVLAV	VLCTLTGIKI	ADYPKAVWQG	AKSMFGAIAI	LILAWLISTV	VGEMHTGDYL
		310	320	330	340	350	360
50	orf26a.pep	STLVAGNIHP	GFLXVILFLL	ASVMAFATGT	SWGTFGIMLP	IAAAMAVKVD	PSLIIPCMSA
	orf26-1	STLVAGNIHP	GFLXVILFLL	ASVMAFATGT	SWGTFGIMLP	IAAAMAVKVD	PSLIIPCMSA
		370	380	390	400	410	420
55	orf26a.pep	VMAGAVCGDH	CSPISDTTIL	SSTGARCNIH	DHVTSQLPYA	LTVAAAAASG	YLALGLTKSA
	orf26-1	VMAGAVCGDH	CSPISDTTIL	SSTGARCNIH	DHVTSQLPYA	LTVAAAAASG	YLALGLTKSA
		430	440	450	460	470	480
60	orf26a.pep	LLFGXGTGIV	LAVLIFLLKD	KKRANA*			
	orf26-1	LLFGXGTGIV	LAVLIFLLKD	KKRANA*			
		490	500				
65	orf26a.pep	LLFGXGTGIV	LAVLIFLLKD	KKRANA*			
	orf26-1	LLFGXGTGIV	LAVLIFLLKD	KKRANA*			

or26-1                    |||||:|||||  
 LLGFGTTGIVLAVLIFLLKDKKRANAX  
                          490                500

# 5 Homology with a predicted ORF from *N. gonorrhoeae*

ORF26 shows 94.8% and 99% identity in 97 and 206 aa overlap at the N-terminus and C-terminus, respectively, with a predicted ORF (ORF26ng) from *N. gonorrhoeae*:

	or26.pep	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILXXVAFVLVGGNPVDGLTHLKDMV	60
10	or26ng	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV	60
	or26.pep	VGLAWSDXDWSLKGPKILVFXILLGIFTSLLTYSGSN	97
15	or26ng	VGLAWADGDWSLKGPKILVFLILLGIFTSLLTYSGSNQAFADWAKRHIKNRCGAKMLTAC	120
		//	
	or26.pep	TSLVFGGTCGVFAVVLCTLGTIKTADYPKA	326
20	or26ng	ASTVSAMIYTGAQASETFSILGAFENTDVNTSLVFGGTCGVLA VVLCTFGTIKTADYPKA	326
	or26.pep	VWQGA KSMFGAIAILILAWLISTVVGEMHTGDYLS TLVAGNIHPGFLPVILFLIASVMAF	386
	or26ng	VWQGA KSMFGAIAILILAWLISTVVGEMHTGDYLS TLVAGNIHPGFLPVILFLIASVMAF	386
25	or26.pep	ATGTSWGTFGIMLP IAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPI SDTTILSSTGAR	446
	or26ng	ATGTSWGTFGIMLP IAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPI SDTTILSSTGAR	446
30	or26.pep	CNHIDHVTSQLPYALTVA AAAASGYLALGLTKSALLGFGTTGIVLAVLIFLLKDKK	502
	or26ng	CNHIDHVTSQLPYALTVA AAAASGYLALGLTKSALLGFGTTGIVLAVLIFLLKDKKRADV	506

The complete length ORF26ng nucleotide sequence <SEQ ID 695> is:

	1	ATGCAGCTGA	TTGACTATTC	ACATTCA TTT	TTCTCGGTTG	TGCCACCCTT
35	51	TTTGGCACTG	GCACTTGCCG	TCATTACCG	CCGCGTACTG	CTGTCTTTAG
	101	GCATCGGTAT	TTTGGTCGGC	GTTGCC TTT	TGGTCGGCGG	CAACCCCGTC
	151	GACGGTCTGA	CACACCTGAA	AGACATGGTC	GTCGGCTTGG	CTTGGGCAGA
	201	CGGCGATTGG	TCGCTGGGCA	AACCAAAAT	CTTGGTTTTC	CTGATACTTT
	251	TGGGCATTTT	CACTTCACTG	CTGACCTACT	CCGGCAGCAA	TCAGGCGTTT
40	301	GCCGACTGGG	CAAAACGGCA	CATTAAAAC	CGGTGCGGCG	CGAAAATGCT
	351	GACCGCCTGC	CTCGTGTTCG	TAACCTTTAT	CGACGACTAT	TTCCACAGCC
	401	TCGCCGTCGG	TGCGATTGCC	CGCCCCGTTA	CCGACAAGTT	TAAAGTTTCC
	451	CGCGCCAAAC	TCGCCATACAT	CCTCGACTCC	ACTGCCTCGC	CCATGTGCGT
	501	GCTGATGCCC	GTTTCAAGCT	GGGGCGCGTC	GATTATCGCC	ACGCTTGCCG
45	551	GATTGCTCGT	TACCTACAAA	ATTACCGAAT	ACACGCCGAT	GGGGACGTTT
	601	GTCGCCATGA	GCCTGATGAA	CTATTACGCG	CTGTTTGCCC	TGATTATGGT
	651	ATTCGTCTGC	GCATGGTTCT	CCTTCGACAT	CGGCTCGATg	gCGCGTTTCG
	701	AACAGGCTGC	GTTGAACGAA	gcccaggacg	aaaccgcccgc	tTCAGACgCT
	751	ACCAAAGGTC	GTGTTTACGC	ATTGATTATT	CCCGTTTTTG	CCTTAATCGC
50	801	CTCAACGGTT	TCCGCCATGA	TCTACACCGG	CGCGCAGGCA	AGCGAAACCT
	851	TCAGCATTTT	GGGGGCATT	GAAAATACCG	ACGTAAACAC	TTCGCTGGTA
	901	TTCGGCGGCA	CTTGCGCGGT	GCTTGCCGTC	GTCCTCTGCA	CGTTCCGGCAC
	951	GATTAAAACC	GCCGATTATC	CCAAAGCCGT	GTGGCAGGGT	GCGAAATCCA
	1001	TGTTCCGGCG	AATCGCCATT	TTAATCCTCG	CCTGGCTCAT	CAGTACGGTT
55	1051	GTCGGCGAAA	TGCACACGGG	CGACTACCTC	TCCACGCTGG	TTGCGGGCAA
	1101	CATCCATCCC	GGCTTCCTGC	CCGTATCCT	CTTCTGCTC	GCCAGCGTGA
	1151	TGGCGTTTGC	CACAGGCACA	AGCTGGGGGA	CGTTCCGGCAT	TATGCTGCCG
	1201	ATTGCCGCGG	CCATGGCGGT	CAAAGTCGAA	CCCGCGCTGA	TTAtcccGTG
	1251	TATGTCCGCA	GTAATGGCGG	GGGCGGTATG	CGGCGACCAC	TGTTCCGCCA
60	1301	TCTCCGACAC	GACCATCCTG	TCGTCCACCG	GCGCGCGCTG	CAACCATATC
	1351	GACCACGTTA	CCTCGCAACT	GCCTTATGCC	CTGACGGTTG	CCGCGCGCGC
	1401	CGCATCGGGC	TACCTCGCAT	TGGGTCTGAC	AAAATCCGCG	CTGTGGGGCT
	1451	TTGGCACGAC	CGGTATTGTA	TTGGCGGTGC	TGATTTTCT	GTTGAAAGAT
	1501	AAAAACGCG	CCGACGTTTG	A		

65 This encodes a protein having amino acid sequence <SEQ ID 696>:

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFLVGGNPV  
 51 DGLTHLKDMV VGLAWADGDW SLGKPKILVF LILGIFTSL LTYSGSNQAF  
 101 ADWAKRHIKN RCGAKMLTAC LVEVTFIDDY FHS LAVGAIA RPVTDKFKVS  
 151 RAKLAYILDS TASPMLVMP VSSWGASTIA TLAGLLVYTK ITEYTPMGTF  
 201 VAMSLMNYIA LFALIMFEVV AWFSFDIGSM ARFEQAALNE AQDETAASDA  
 251 TKGRVYALII PVLALIASTV SAMIYTGAQA SETFSILGAF ENTDVNTSLV  
 301 FGGTCGVLA VLTCTGTIKT ADYPKAVWQG AKSMFGAIAI LILAWLISTV  
 351 VGEMHTGDYL STL VAGNIHP GFLPVILFLL ASVMAFATGT SWGTFGIMLP  
 401 IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARNHI  
 451 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGTTGIV LAVLIFLLKD  
 501 KKRADV\*

ORF26ng and ORF26-1 show 98.4% identity in 505 aa overlap:

15	orf26-1.pep	10 20 30 40 50 60	MQLIDYSHSEFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV
	orf26ng	10 20 30 40 50 60	MQLIDYSHSEFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV
20	orf26-1.pep	70 80 90 100 110 120	VGLAWSGDGWSLGKPKILVFLILLGIFTSLTYSGSNQAFADWAKRHIKNRRGAKMLTAC
	orf26ng	70 80 90 100 110 120	VGLAWADGDWSLGKPKILVFLILLGIFTSLTYSGSNQAFADWAKRHIKNRCGAKMLTAC
25	orf26-1.pep	130 140 150 160 170 180	LVEVTFIDDYFHS LAVGAIAARPVTDKFKVSRTKLAYILDSTAAPMCVLMPVSSWGASTIA
	orf26ng	130 140 150 160 170 180	LVEVTFIDDYFHS LAVGAIAARPVTDKFKVSRAKLAYILDSTAAPMCVLMPVSSWGASTIA
30	orf26-1.pep	190 200 210 220 230 240	TLAGLLVYTKITEYTPMGTFVAMSLMNYIALFALIMFEVVAVWFSFDIGSMARFEQAALNE
	orf26ng	190 200 210 220 230 240	TLAGLLVYTKITEYTPMGTFVAMSLMNYIALFALIMFEVVAVWFSFDIGSMARFEQAALNE
35	orf26-1.pep	250 260 270 280 290 300	AHDETAVS DATKGRVYALII PVLALIASTVSAMIYTGAQASETFSILGAFENTDVNTSLV
	orf26ng	250 260 270 280 290 300	AQDETAAS DATKGRVYALII PVLALIASTVSAMIYTGAQASETFSILGAFENTDVNTSLV
45	orf26-1.pep	310 320 330 340 350 360	FGGTCGVLA VLTCTGTIKTADYPKAVWQGA KSMFGAIAI LILAWLISTVVGEMHTGDYL
	orf26ng	310 320 330 340 350 360	FGGTCGVLA VLTCTGTIKTADYPKAVWQGA KSMFGAIAI LILAWLISTVVGEMHTGDYL
50	orf26-1.pep	370 380 390 400 410 420	STLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLP IAAAMAVKVEPALIIPCMSA
	orf26ng	370 380 390 400 410 420	STLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLP IAAAMAVKVEPALIIPCMSA
55	orf26-1.pep	430 440 450 460 470 480	VMAGAVCGDH CSPISDTTILSSTGARNHIDHVTSQLPYALTVA AAAASGYLALGLTKSA
	orf26ng	430 440 450 460 470 480	VMAGAVCGDH CSPISDTTILSSTGARNHIDHVTSQLPYALTVA AAAASGYLALGLTKSA
60	orf26-1.pep	490 500	LLGFGTTGIVLAVLIFLLKDKKRANAX
	orf26ng	490 500	LLGFGTTGIVLAVLIFLLKDKKRADVX
65			

In addition, ORF26 ng shows significant homology to a hypothetical *H. influenzae* protein:

-400-

sp|P44263|YF86\_HAEIN HYPOTHETICAL PROTEIN HI1586 >gi|1074850|pir||C64037  
 hypothetical  
 protein HI1586 - Haemophilus influenzae (strain Rd KW20) >gi|1574427 (U32832) H.  
 influenzae predicted coding region HI1586 [Haemophilus influenzae] Length = 519  
 Score = 538 bits (1370), Expect = e-152  
 Identities = 280/507 (55%), Positives = 346/507 (68%), Gaps = 7/507 (1%)

Query: 1 MQLIDYSHSFFSVVPFLALALAVITRRXXXXXXXXXXXXXAFVGGNPVDGLTHLKDMV 60  
 M+LID+S S +S+VP LA+ LA+ TRR L +L V  
 Sbjct: 14 MELIDFSSSVWSIVPALLAIIAIAIATRRVLVLSAGIIGSLMLSDWQIGSAFNYLVKNV 73

Query: 61 VGLAWADGDWSLGKPKILVFLILLGIFTSLLTYSGSNQAFADWAKRHIKNRCGAKMLTAC 120  
 V L +ADG+ + I++FL+LLG+ T+LLT SGSN+AFA+WA+ IK R GAK+L A  
 Sbjct: 74 VSLVYADGEIN-SNMNIVLFLLLLGVLTAALLTVSGSNRAFAEWAQSRIGRRGAKLLAAS 132

Query: 121 LVFVTFIDDFHSLAVGAIARPVTDKFKVSRKLAYILDSTASPMCVLMPVSSWGASIIA 180  
 LVFVTFIDDFHSLAVGAIARPVTD+FKVSRKLAYILDSTA+PMCV+MPVSSWGA II  
 Sbjct: 133 LVFVTFIDDFHSLAVGAIARPVTDKFKVSRKLAYILDSTAAPMCVMPVSSWGAYIIT 192

Query: 181 TLAGLLVITYKITEYTPMGTFVAMSLMNYALFALIMVFWVAFSFDIGSMARFEQAALNE 240  
 + GLL TY ITEYTP+G FVAMS MN+YA+F++IMVF VA+FSFDI SM R E+ AL  
 Sbjct: 193 LIGLLATYSITEYTPIGAFVAMSSMNFYAIIFSIIIMVFFVAYFSFDIASMVRHEKLALKN 252

Query: 241 AQDETAASDATKGRVYALIIPVLALIASTVSAMIYTGAQA----SETFSILGAFENTDVN 296  
 +D+ TKG+V LI+P+L LI +TVS MIYTGA+A + FS+LG FENT V  
 Sbjct: 253 TEDQLEETGTGKQVRNLILPILVLIATVSMMIYTGAELAADGKVFVSVLGTFFENTVVG 312

Query: 297 TSLVFGGTCGVL--AVVLCTFGTIKTADYPKAVWQAKSMFGXXXXXXXXXXXXSTVVGEM 354  
 TSLV GG C ++ +++ + +Y ++ G KSM G + +VG+M  
 Sbjct: 313 TSLVVGGFCSIIISTLLIILDRQVSVPEYVRSWIVGIKSMGAIAILFFAWTINKIVGDM 372

Query: 355 HTGDYLSLTVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLPAAAMAVKVEPALI 414  
 TG YLS+LV+GNI FLPVIL+L + MAF+TGTWGTFGIMLPAAAMA P L+  
 Sbjct: 373 QTGKYLSSLVSGNIPMQFLPVILFVLGAAMAFSTGTSWGTFGIMLPAAAMAANAPELL 432

Query: 415 IPCMSAVMAGAVCGDHCSPISTTILSSTGARNHIDHVT SQXXXXXXXXXXXXXXXXXXXX 474  
 +PC+SAVMAGAVCGDHCS+SDTTILSSTGA+CNHIDHVT+Q  
 Sbjct: 433 LPCLSAVMAGAVCGDHCSFVSDTTILSSTGAKCNHIDHVTQLPYAATVATATSIGYIVV 492

Query: 475 XXXKSALLGFGTTGIVLAVLIFLLKDK 501  
 S L GF T + L V+IF +K +  
 Sbjct: 493 GFTYSLAGFAATAVSLIVIIFAVKKR 519

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*,  
 and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 83

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 697>:

1 ..AAGCAATGGT ATGCCGACGN .AGTATCAAG ACGGAAATGG TTATGGTCAA  
 51 CGATGAGCCT GCCAAATTC TGACTTGGGA TGAAAGCGGC CGATTACTCT  
 101 CGGAAGTGT TATCCGCCAC CATCAACGCA ACGGGGTGGT TTTGGAGTGG  
 151 TATGAAGATG GTTCTAAAAA GAGCGAAGT. GTTTATCAGG ATGACAAGTT  
 201 GGTCAGGAAA ACCCAGTGGG ATAAGGATGG TTATTTAATC GAACCTGA

This corresponds to the amino acid sequence <SEQ ID 698; ORF27>:

1 ..KQWYADXSIK TEMVMVNDEP AKILTWDESG RLLSELSIRH HQRNGVVLEW  
 51 YEDGSKKSEX VYQDDKLVRK TQWDKGYLI EP\*

Further work revealed the complete nucleotide sequence <SEQ ID 699>:

1 ATGAAAAAAT TATCTCGGAT TGTATTTTCA ACTGTCCTGT TGGGTTTTTC  
 51 GGCCGCTTTG CCGGCGCAGA CCTATTCTGT TTATTTTAAT CAGAACGGAA  
 101 AGCTGACGGC GACGATGCTT TCTGCCGCTT ATATCAGGCA ATATAGTGTG  
 151 GTGGCGGTA TTGCGCACGC GCAGGATTTT TATTATCCGT CGATGAAGAA

10

201	ATATTCTGAA	CCTTATATCG	TTGCTTCAAC	GCAAAATCAA	TCTTTTGTGC
251	CTACCCTGCA	AAACGGTATG	TTGATTTTGT	GGCATTTTAA	TGGTCAGAAA
301	AAAATGGCGG	GGGCGTTCAG	CAAGGGTAAG	CCGGACGGGG	AGTGGGTCAA
351	CTGGTATCCG	AACGGTAAAC	AATCTGCCGT	TATGCCTTAT	AAAAATGGCT
401	TGAGTGAGG	TACGGGATAC	CGCTATTACC	GTAACGGCGG	CAAGGAAAGC
451	GAAATCCAGT	TTAAGCAAAA	TAAGGCAAAC	GGCGTATGGA	AGCAATGGTA
501	TGCCGACGGC	AGTATCAAGA	CGGAAATGGT	TATGGTCAAC	GATGAGCCTG
551	CCAAAATTCT	GACTTGGGGT	GAAAGCGGGC	GATTACTCTC	GGAACTGTCT
601	ATCCGCCACC	ATCAACGCAA	CGGGGTGGTT	TGTGAGTGGT	ATGAAGATGG
651	TTCTAAAAAG	AGCGAAGCTG	TTTATCAGGA	TGACAAGTTG	GTACAGGAAAA
701	CCGAGTGGGA	TAAGGATGGT	TATTTAATCG	AACCCCTGA	

15

1	MKKLSRIVES	TVLLGFSAL	PAQTYSVYFN	QNGKLTATMS	SAAYIROQSV
51	<u>VAGIAHAQDF</u>	YYPMSKKYSE	PYIVASTQIK	SFVPTLQNGM	LILWHFNGQK
101	KMAGGFSK GK	PDGEWVNWYP	NGKKSAMVPM	KNGLSEGTGY	RYRRNGGKES
151	EIQGFQNKAN	GVWKQYIADG	SIKTEMVMVN	DEKALITWD	ESGRLLSELS
201	IRHHORNGVV	LWEYEDGSKK	SEAVYQDDKL	VRKTQWDKDG	YLIEP*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

20 ORF27 shows 91.5% identity over a 82aa overlap with an ORF (ORF27a) from strain A of *N. meningitidis*:

[illegible]

The complete length ORF27a nucleotide sequence <SEQ ID 701> is:

35	1	ATGAAAAAAT	TATCTCGGAT	TGTATTTTCA	ACTGTCCTGT	TGGGTTTTTC
	51	GGCCGCTTTG	CCGGCGCAGA	NCTATTCTGT	TTATTTTAAAT	CAGAACGGGA
	101	AACCTGACGGC	GACGNTGTCT	TCTGCCGCNT	ATATCAGGCA	ATATAGTGTG
	151	CGGGAGGGTA	TTGCGCACGC	GCAGGANTTT	TANTATCCGT	CGATGAAGAA
	201	ATATTCCGAA	CCTTATATCG	TTGCTTCAAC	GCAAAATCAA	TCTTTTGTGC
40	251	CTACCTGCA	AAACGGTATG	TTGATTTTGT	GGCATTTTAA	NGGTGAGAAA
	301	AAAATGGCNG	GGGGCTTCAG	CAAGGGTAAG	CCGGACGGGG	AGTGGGTCAA
	351	CTGGTATCCG	AACGGTA AAA	AATCTGCCGT	TATGCCTTAT	AAAAATGGTT
	401	TGAGTGAAGG	TACGGGGTNN	CGCTATTACC	GTACCGGGCG	CAAGGAAAGC
	451	GAAATCCAGT	TTAAACAGAA	TAAGGCAAAAC	GGCGTATGGA	AGCAATGGTA
45	501	TCCGACGGC	ATAATCAAAA	CGGAAATGGT	TATGGTCAAT	GATGAGCCCTG
	551	CCAAAATTCT	GACATGGGAT	GAAAGCGGTC	GATTACTCTC	GGAAGTGTCT
	601	ATCCATCATC	ATNAACGTAA	TGGAGTAGTC	TTAGAGTGGT	ATGAAGATGG
	651	TTCTAAAAAG	ANTGAAGCTG	TTTATCAGGA	TGATAAGTTG	GTCAGGAAAA
	701	CCAGATGGGA	TAANGATGGT	TATTTAATCG	AACCTCGA	

This encodes a protein having amino acid sequence <SEQ ID 702>:

50	1	<u>MKKLSRIVFS</u>	<u>TVLLGFSAA</u>	<u>PAQXYSVYFN</u>	<u>QNGKLTATXS</u>	<u>SAAYIRQYSV</u>
	51	<u>AEGIAHAQXF</u>	<u>XYPSMKKYSE</u>	<u>PYIVASTQIK</u>	<u>SFVPTLQNGM</u>	<u>LILWHFXGQK</u>
	101	<u>KMAGGFSGKG</u>	<u>PDGEWVNWYP</u>	<u>NGKKSAVMPY</u>	<u>KNGLSEGTGX</u>	<u>RYRRNGGKES</u>
	151	<u>EIQFKQNKAN</u>	<u>GVWKQWYADG</u>	<u>NIKTEMVMVN</u>	<u>DEPAKLTWID</u>	<u>ESGRLLSELS</u>
	201	<u>IHHHXRNGVU</u>	<u>LEWYEDGSKK</u>	<u>XEAVYODDKL</u>	<u>VRKTQWDXDG</u>	<u>YLIEP*</u>

55 ORF27a and ORF27-1 show 94.7% identity in 245 aa overlap:

orf27a.pep MKKLSRIVFSTVLLGFSALPAQXYSVYFNQNGKLTATXSSAAYIROYSVAEGIAHAQXF

-402-

	or27-1	 MKKLSRIVFSTVLLGFSAALPAQTYSVYFNQNGKLTATMSSAAYIRQYSVVAGIAHAQDF 10 20 30 40 50 60
5	or27a.pep	70 80 90 100 110 120 XYPMSKKYSEFYIVASTQIKSFVPTLQNGMLILWHFXGQKKMAGGFSKGPDEGEVWNWYP 
10	or27-1	70 80 90 100 110 120 YYPMSKKYSEFYIVASTQIKSFVPTLQNGMLILWHFNGQKKMAGGFSKGPDEGEVWNWYP 
	or27a.pep	130 140 150 160 170 180 NGKKSAMVPYKNGLSEGTGXRYRNGGKESEIQFKQNKANGVWKQWYADGNIKTEMVMVN 
15	or27-1	130 140 150 160 170 180 NGKKSAMVPYKNGLSEGTGYRYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN 
	or27a.pep	190 200 210 220 230 240 DEPAKILTWDSEGRLLSELSIRHHXRNQGVVLEWYEDGSKKXEAQYQDDKLVRKTQWDKDG 
20	or27-1	190 200 210 220 230 240 DEPAKILTWDSEGRLLSELSIRHHQRNQGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDG 190 200 210 220 230 240
25	or27a.pep	YLIEPX 
	or27-1	YLIEPX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF27 shows 96.3% identity over 82 aa overlap with a predicted ORF (ORF27ng) from

30 *N.gonorrhoeae*:

	or27.pep	KQWYADXS IKTEMVMVNDEPAKILTWDSEG 30
	or27ng	LSEGTGYRYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVNDEPAKILTWDSEG 193
35	or27.pep	RLLSELSIRHHQRNQGVVLEWYEDGSKKSEXVYQDDKLVRKTQWDKDG YLIEP 82
	or27ng	RLLSELSIRHHQRNQGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDG YLIEP 245

The complete length ORF27ng nucleotide sequence &lt;SEQ ID 703&gt; is:

40	1	ATGAAGAAAT TATCTCGGAT TGTATTTTCA ATCGTACTGT TGGGTTTTTC
	51	GGCCGCTTTG CCGGCGCAGA CCTATTCTGT TTATTTTAAT CAGAACGGGA
	101	AACTGACGGC GACGATGTCT TCTGCCGCTT ATATCAGGCA ATATAGTGTG
	151	GCGGCGGGTA TCGCACACGC GCAGGATTTT TATTATCCGT CGATGAAGAA
	201	ATATTCGGAA CCTATATCG TTGCTTCAAC GCAATCAAA TCTTTTGTGC
45	251	CTACCCTGCA AAACGGTATG TTGATTTTGT GGCATTTTAA TGGTCAGAAA
	301	AAAATGGCGG GGGCTTTCAG CAAGGGTAAG CCGGACGGGG AATGGGTCAA
	351	CTGGTATCCG AACGGTAAAA AATCTGCGGT TATGCCTTAT AAAATGGCT
	401	TGAGTGAGGG TACGGGATAC CGTTATTACC GTAACGGCGG CAAGGAAAGC
	451	GAAATCCAGT TTAAGCAAAA TAAGGCGAAC GGCATGATGA AGCAATGGTA
50	501	TGCCGATGGA AGTATCAAGA CGGAAATGGT TATGGTCAAC GATGAGCCTG
	551	CCAAAATTCT GACTTGGGAT GAAAGCGGCC GATTACTTTC GGAAGTGTCT
	601	ATCCGCCACC ATAAACGCAA CGGGGTGGTT TTGGAGTGGT ATGAAGATGG
	651	TTCTAAAAAG AGCGAGGCTG TTTATCAGGA TGACAAGTTG GTCAGGAAAA
	701	CCCAATGGGA TAAGGATGGT TATTTAATCG AACCTGA

This encodes a protein having amino acid sequence &lt;SEQ ID 704&gt;:

55	1	MKKLSRIVFS IVLLGFSAAL PAQTYSVYFN QNGKLTATMS SAAYIRQYSV
	51	AAGIAHAQDF YYPMSKKYSE FYIVASTQIK SFVPTLQNGM LILWHFNGQK
	101	KMAGGFSKGP PDGEVWNWYP NGKKSAMVPY KNGLSEGTGY RYRNGGKES
	151	EIQFKQNKAN GVWKQWYADG SIKTEMVMVN DEPAKILTWD ESGRLLSELS
	201	IRHHKRNGVV LEWYEDGSKK SEAVYQDDKL VRKTQWDKDG YLIEP*

60 ORF27ng and ORF27-1 show 98.8% identity in 245 aa overlap:

	or27-1.pep	10 20 30 40 50 60 MKKLSRIVFSTVLLGFSAALPAQTYSVYFNQNGKLTATMSSAAYIRQYSVVAGIAHAQDF
--	------------	---

	orf27ng	MKKLSRIVFSIVLLGFSAALPAQTYSVYFNQNGKLTATMSSAAYIRQYSVAAGIAHAQDF	10	20	30	40	50	60
5	orf27-1.pep	YYPSPMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFNGQKKMAGGFSKGPGEWVNWYP	70	80	90	100	110	120
	orf27ng	YYPSPMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFNGQKKMAGGFSKGPGEWVNWYP	70	80	90	100	110	120
10	orf27-1.pep	NGKKSAVMPYKNGLSEGTGYRYYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN	130	140	150	160	170	180
	orf27ng	NGKKSAVMPYKNGLSEGTGYRYYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN	130	140	150	160	170	180
15	orf27-1.pep	DEPAKILTWDESGRLLSELSIRHHQRNGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDG	190	200	210	220	230	240
20	orf27ng	DEPAKILTWDESGRLLSELSIRHHQRNGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDG	190	200	210	220	230	240
25	orf27-1.pep	YLIEPX						
	orf27ng	YLIEPX						

30

35

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 705>:

40

45

50

1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WGFYGTGTHX LSGFYWHAHE  
51 MWGYAGLVV IAFLLTAVAT WTGQPTRGG VLVGLTIFWL AARIAAFIPG  
101 WGASASGILG TLEFFWYGAVC MALPVIRSON QRNYVAVFAL FVLGGTHAAF



151 HVQLHNGNLG GLLSGLQSGL VM

Further work revealed the complete nucleotide sequence <SEQ ID 707>:

```

      1 ATGAAATTTA CCAAGCACCC CGTCTGGGCA ATGGCGTTCC GCCCATTTTA
      51 TTCGCTGGCG GCTCTGTACG GCGCATTGTC CGTATTGCTG TGGGGTTTCG
5      101 GCTACACGGG AACGCACGAG CTGTCCGGTT TCTATTGGCA CGCGCATGAG
      151 ATGATTTGGG GTTATGCCGG ACTGGTCGTC ATCGCCTTCC TGCTGACCGC
      201 CGTCGCCACT TGGACGGGGC AGCCGCCCAC GCGGGGCGGC GTTCTGGTCG
      251 GCTTGACTAT CTTTGGCTG GCTGCGCGGA TTGCCGCTT TATCCCGGGT
      301 TGGGGTGCGT CGGCAAGCGG CATACTCGGT ACGCTGTTT TCTGGTACGG
10     351 CGCGTGTGCG ATGGCTTTGC CCGTTATCCG TTCGCAGAAT CAACGCAACT
      401 ATGTTGCCGT GTTCGCGCTG TTCGTCTTGG GCGGCACGCA TGCGGCGTTC
      451 CACGTCCAGC TGCACAACGG CAACCTAGGC GGACTCTTGA GCGGATTGCA
      501 GTCGGGCTTG GTGATGGTGT CGGGTTTTAT CGGTCTGATT GGTACGCGGA
      551 TTATTTCTGT TTTTACGTCC AAACGCTTGA ATGTGCCGCA GATTCCCACT
15     601 CCGAAATGGG TGGCGCAGGC TTCGCTGTGG CTGCCCATGC TGAATGCCAT
      651 GCTGATGGCG CACGGTGTGT TGGCTTGGCT GTCTGCCGTT TTTGCCCTTG
      701 CGGCAGGTGT GATTTTTACC GTGCAGGTGT ACCGCTGGTG GTATAAACCC
      751 GTGTTGAAAG AGCCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC
      801 CGGATTTGGG CTGATTGCGG TCGGCGCGTC TTATTTCAAA CCCGCTTTC
20     851 TCAATCTGGG TGTGCATCTG ATCGGGGTCG GCGGTATCGG CGTGCTGACT
      901 TTGGGCATGA TGGCGCGTAC CGCGCTTGGT CATAACGGCA ATCCGATTTA
      951 TCCGCCGCCC AAAGCCGTT CCGTTGCGTT TTGGCTGATG ATGGCGGCAA
     1001 CCGCGTCCG TATGGTTGCC GTATTTTCTT CCGGCACTGC CTACACGCAC
     1051 AGCATCCGCA CCTCTTCGGT TTTGTTTGCA CTCGCGCTTT TGGTGTATGC
25     1101 GTGGAAGTAT ATTCCTTGGC TGATTCTGTC GCGTTCGGAC GGCAGGCCCG
     1151 GTTGA
  
```

This corresponds to the amino acid sequence <SEQ ID 708; ORF47-1>:

```

      1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WFGYTGTHE LSGFYWHAHE
      51 MIWGYAGLVV IAFLLTAVAT WTGQPPTRGV VLVGLTIFWL AARIAAFIPG
30     101 WGASASGILG TLEFFWYAVC MALPVIRSON QRNYVAVFAL FVLGGTHAAF
      151 HVQLHNGNLG GLLSGLQSGL VMVSGFIGLI GTRIISFFTS KRLNVPQIPS
      201 PKWVAQASLW LPMLTAMLMA HGVLAWSAV FAFAAGVIFT VQVYRWYKYP
      251 VLKEPMLWIL FAGYLFGLG LIAVGASYFK PAFLNLGVHL IGVGGIGVLT
      301 LGM MARTALG HTGNPIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
35     351 SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRPG*
  
```

Computer analysis of this amino acid sequence predicts a leader peptide and also gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF47 shows 99.4% identity over a 172aa overlap with an ORF (ORF47a) from strain A of *N.*

40 *meningitidis*:

```

      10      20      30      40      50      60
orf47.pep MKFTKHPVWMAFRPFYSLAALYGALSVLLWFGYTGTHXLSGFYWHAHEMIWGYAGLVV
      10      20      30      40      50      60
45 orf47a  MKFTKHPVWMAFRPFYSLAALYGALSVLLWFGYTGTHLSGFYWHAHEMIWGYAGLVV
      10      20      30      40      50      60
      70      80      90      100     110     120
orf47.pep IAFLLTAVATWTGQPPTRGVVLVGLTIFWLAARIAAFIPGWGASASGILGLTLEFFWYAVC
      10      20      30      40      50      60
50 orf47a  IAFLLTAVATWTGQPPTRGVVLVGLTIFWLAARIAAFIPGWGASASGILGLTLEFFWYAVC
      70      80      90      100     110     120
      130     140     150     160     170
orf47.pep MALPVIRSONQRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGLVM
      10      20      30      40      50      60
55 orf47a  MALPVIRSONQRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGLVMVSGFIGLI
      130     140     150     160     170     180
      190     200     210     220     230     240
60 orf47a  GTRIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMAHGVMPWLSAAFAFAAGVIFT
  
```

The complete length ORF47a nucleotide sequence <SEQ ID 709> is:

```

1  ATGAAATTTA CCAAGCACCC CGTTTGGGCA ATGGCGTTCC GCCCGTTTTA
51  TTCCTGGCG GCTCTGTACG GCGCATTGTC CGTATTGCTG TGGGGTTTCG
101 GCTACACGGG AACGCACGAG CTGTCCGGTT TCTATTGGCA CGCGCATGAG
5  151  ATGATTGGG GTTATGCCGG ACTGGTCGTC ATCGCCTTCC TGCTGACCGC
201  CGTCGCCACT TGGACGGGCG AGCCGCCAC GCGGGGCGGC GTTCTGGTCG
251  GCTTGACTAT CTTTGGCTG GCTGCGCGGA TTGCCGCTT TATCCGGGT
301  TGGGGTGCCT CGGCAAGCGG CATACTCGGT ACGCTGTTT TCTGGTACGG
351  CGCGGTGTGC ATGGCTTTGC CCGTTATCCG TTCGCAGAAT CAACGCAATT
10  401  ATGTTGCCGT GTTCGCGCTG TTCGTCTTGG GCGGTACGCA CGCGCGCTTC
451  CACGTCCAGC TGCACAACGG CAACCTAGGC GGACTCTTGA GCGGATTGCA
501  GTCGGGCTTG GTGATGGTGT CGGGTTTAT CCGTCTGATT GGTACGCGGA
551  TTATTTTCGT TTTTACGTCC AAACGGTTGA ATGTGCCGCA GATTCCAGT
601  CCGAAATGGG TGGCGCAGGC TTCGCTGTGG CTGCCCATGC TGACCGCCAT
15  651  GCTGATGGCG CACGCGCTGA TGCCTTGGCT GTCGGCGGCT TTCGCGTTTG
701  CGGCAGGTGT GATTTTTACC GTGCAGGTGT ACCGCTGGTG GTATAAGCCT
751  GTGTTGAAAG AGCCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC
801  CGGATTGGGG CTGATTGCGG TCGGCGCGTC TTATTTCAA CCCGCTTTC
851  TCAATCTGGG TGTGCTCTG ATCGGGGTCG GCGGTATCGG CGTGCTGACT
20  901  TTGGGCATGA TGGCGCGTAC CGCGCTCGGT CACACGGGCA ATCCGATTTA
951  TCCGCCGCC AAAGCCGTT CCGTTGCGTT TTGGCTGATG ATGGCGGCAA
1001 CCGCCGTCG TATGGTTGCC GTATTTTCTT CCGGCACTGC CTACACGCAC
1051 AGCATACGCA CCTCTTCGGT TTGTTTGCA CTCGCGCTT TGGTGTATGC
1101 GTGGAAGTAT ATTCCTTGGC TGATTCTGCC GCGTTCGGAC GGCAGGCCCG
25  1151 GTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 710>:

```

1  MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WFGYGTGTHE LSGFYWHAHE
51  MIWGYAGLVV IAFLLTAVAT WTGPPTRGV VLVGLTIFWL AARIAAFIPG
101  WGASASGILG TLFFWYGAVC MALPVIRSON QRNYVAVFAL FVLGGTHAAF
30  151  HVQLHNGNLG GLLSGLQSLG VMVSGFIGLI GTRIISEFTS KRLNVPQIPS
201  PKWVAQASLW LPMLTAMLM MAHVMPWLSAA FAFAGVIFT VQVYRWYKYP
251  VLKEPMLWIL FAGYLETGLG LIAVGASYFK PAFLNLGVHL IGVGGIGVLT
301  LGMMARTALG HTGNPIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
351  SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRPG*

```

ORF47a and ORF47-1 show 99.2% identity in 384 aa overlap:

```

10      20      30      40      50      60
orf47a.pep  MKFTKHPVWAMAFRPFYSLAALYGALSVLLWFGYGTGTHELSGFYWHAHEMIWGYAGLVV
40  orf47-1  MKFTKHPVWAMAFRPFYSLAALYGALSVLLWFGYGTGTHELSGFYWHAHEMIWGYAGLVV
      10      20      30      40      50      60
70      80      90      100     110     120
orf47a.pep  IAFLLTAVATWTGPPTRGVVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
45  orf47-1  IAFLLTAVATWTGPPTRGVVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
      70      80      90      100     110     120
130     140     150     160     170     180
orf47a.pep  MALPVIRSONQRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGLMVMSGFIGLI
50  orf47-1  MALPVIRSONQRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGLMVMSGFIGLI
      130     140     150     160     170     180
190     200     210     220     230     240
orf47a.pep  GTRIISEFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMAGVMPWLSAAFAFAAGVIFT
55  orf47-1  GTRIISEFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMAGVMPWLSAAFAFAAGVIFT
      190     200     210     220     230     240
250     260     270     280     290     300
orf47a.pep  VQVYRWYKYPVLKEPMLWILFAGYLETGLGLIAVGASYFKPAFLNLGVHLIGVGGIGVLT
60  orf47-1  VQVYRWYKYPVLKEPMLWILFAGYLETGLGLIAVGASYFKPAFLNLGVHLIGVGGIGVLT
      250     260     270     280     290     300
310     320     330     340     350     360
65

```

-406-

```

orf47a.pep    LGMMARTALGHTGNPIYPPPKAVPVAFWLMAATAVRMVAVFSSGTAYTHSIRTSSVLFA
               |||
orf47-1       LGMMARTALGHTGNPIYPPPKAVPVAFWLMAATAVRMVAVFSSGTAYTHSIRTSSVLFA
               |||
               310      320      330      340      350      360

               370      380
orf47a.pep    LALLVYAWKYIPWLIRPRSDGRPGX
               |||
orf47-1       LALLVYAWKYIPWLIRPRSDGRPGX
               |||
               370      380

```

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF47 shows 97.1% identity over 172 aa overlap with a predicted ORF (ORF47ng) from *N.gonorrhoeae*:

```

15  ORF47      MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFGYTGTHELSGFYWHAHEMIWGYAGLVV    60
      |||
      ORF47ng  MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFGYTGTHELSGFYWHAHEMIWGYAGLVV    60

20  ORF47      IAFLLTAVATWTGQPPTRGGLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC    120
      |||
      ORF47ng  IAFLLTAVATWTGQPPTRGGLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAVC    120

      ORF47      MALPVIRSQNRNYVAVFALFVLGGTHAAHFVQLHNGNLGGLLSGLQSGSLVM          172
      |||
25  ORF47ng    MALPVIRSQNRNYVAVFAIFVLGGTHAAHFVQLHNGNLGGLLSGLQSGSLVMVWGFIGLI    180

```

The ORF47ng nucleotide sequence <SEQ ID 711> is predicted to encode a protein comprising amino acid sequence <SEQ ID 712>:

```

1  MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WGFGYTGTHE LSGFYWHAHE
30  MIWGYAGLVV IAFLLTAVAT WTGQPPTRGG VLVGLTAFWL AARIAAFIPG
101 WGAAASGILG TLFFWYGAVC MALPVIRSQN RRNYVAVFAI FVLGGTHAAF
151 HVQLHNGNLG GLLSGLQSGL VMVWGFIGLI GMKIISFFTS KRLKLPQIPS
201 PKWVAHASLW LPMLNAILMA HRVMPWLSAA FPFAAGVIFT VOQYAGGITP
251 IEETSCGSVA GICYRLGNSS G

```

The predicted leader peptide and transmembrane domains are identical (except for an Ile/Ala substitution at residue 87 and an Leu/Ile substitution at position 140) to sequences in the meningococcal protein (see also *Pseudomonas stutzeri* orf396, accession number e246540):

```

40  TM segments in ORF47ng
      INTEGRAL    Likelihood = -5.63    Transmembrane    52 - 68
      INTEGRAL    Likelihood = -3.88    Transmembrane    169 - 185
      INTEGRAL    Likelihood = -3.08    Transmembrane    82 - 98
      INTEGRAL    Likelihood = -1.91    Transmembrane    134 - 150
      INTEGRAL    Likelihood = -1.44    Transmembrane    107 - 123
      INTEGRAL    Likelihood = -1.38    Transmembrane    227 - 243

```

Further work revealed the complete gonococcal DNA sequence <SEQ ID 713>:

```

45  1  ATGAAATTTA CCAAACATCC CGTCTGGGCA ATGGCGTTCC GCCCGTTTTA
      51  TTCACTGGCG GCACTGTACG GCGCATTGTC CGTATTGCTG TGGGGTTTCG
101  GCTACACGGG AACGCACGAG CTGTCCGGTT TCTATTGGCA CGCGCATGAG
      151  ATGATTGGGG GTTATGCCGG TCTCGTCGTC ATCGCCTTCC TGCTGACCGC
201  CGTCGCCACT TGGACGGGAC AGCCGCCAC GAGGGGCGGC GTTCTGGTCG
50  251  GCTTGACCGC CTTTGGCTG GCTGCGCGGA TTGCCGCCTT TATCCCGGGT
      301  TGGGGTGC GGCAAGCGG CATACTCGGT ACGCTGTTT TCTGGTACGG
      351  CGCGGTGTGC ATGGCTTTGC CCGTTATCCG TtcgCAAAAC CGGCGCAACT
      401  ATGtcgCCGT ATTCGCAATA TTTGTGCTGG GCGGTACGCA TGCGgcgTTC
      451  CACGtccAgc tGCACAACGG CAACCTAGGC GGACTCTTGA GCGGATTGCA
55  501  GTCGGGCGCT GTTATGGTGT CGGGCTTTAT CGGCCTGATT GGGATGAGGA
      551  TTATTTCTGT TTTTACGTCC AAACGGTTGA ACGTGCCGCA GATTCCCACT
      601  CCGAAATGGG TGGCGCAGGC TTCGCTGTGG CTACCCATGC TGACCGCCAT

```

-407-

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

```

651 ACTGATGGCG CACGGCGTGA TGCCTTGGCT GTCGGCGGCT TTCGCGTTTG
701 CGGCGGGCGT GATTTTACC GTACAGGTGT ACCGCTGGTG GTATAAACCC
751 GTATTGAAAG AACCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC
801 CGGATTGGGG CTGATTGCGG TCGGCGCGTC TTATTCAAA CCTGCCTTCC
851 TCAATCTGGG CGTACATCTG ATCGGGGTCG GCGGTATCGG CGTGCTGACT
901 TTGGGCATGA TGGCGCGTAC CGCGCTCGGT CATACGGGCA ATTCGATTTA
951 TCCGCCGCC AAAGCCGTTT CCGTTGCGTT TTGGCTGATG ATGGCGGCAA
1001 CCGCCGTCCG TATGGTTGCC GTATTTTCTT CCGGCACTGC CTACACGCAC
1051 AGCATCCGCA CGTCTTCGGT TTTGTTTGCA CTCGCGCTGC TGGTGTATGC
1101 GTGGAAATAC ATTCCGTGGC TGATCCGTCC GCGTTCGGAC GGCAGGCCCC
1151 GTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 714; ORF47ng-1>:

15  
20  
25  
30  
35  
40  
45  
50  
55  
60

```

1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WFGYTGTHE LSGFYWHAHE
51 MIWGYAGLVV IAFLLTAVAT WTGQPPTRGV VLVGLTAEWL AARIAAFIPG
101 WGAASGILG TLEFWYGAVC MALPVIRSON RRRNYVAVFAI FVLGGTHAAF
151 HVQLHNGNLG GLLSGLQSLG VMVSGFIGLI GMRIISFSTS KRLNVPQIPS
201 PKWVAQASLW LPMLTAILMA HGVMPWLSAA FAFAGVIFT VQVYRWYKYP
251 VLKEPMLWIL FAGYLFTGLG LIAVGASYFK PAFLNLGVHL IGVGIGVLT
301 LGMMARTALG HTGNSIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
351 SIRTSSVLEA LALLVYAWKY IPWLIRPRSD GRPG*

```

ORF47ng-1 and ORF47-1 show 97.4% identity in 384 aa overlap:

25  
30  
35  
40  
45  
50  
55  
60

```

      10      20      30      40      50      60
orf47-1.pep MKFTKHPVWAMAFRPFYSLAALYGALSVLLWFGYTGTHELSGFYWHAHEMIWGYAGLVV
orf47ng-1    MKFTKHPVWAMAFRPFYSLAALYGALSVLLWFGYTGTHELSGFYWHAHEMIWGYAGLVV
      10      20      30      40      50      60

      70      80      90     100     110     120
orf47-1.pep IAFLLTAVATWTGQPPTRGVVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
orf47ng-1    IAFLLTAVATWTGQPPTRGVVLVGLTAFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
      70      80      90     100     110     120

      130     140     150     160     170     180
orf47-1.pep MALPVIRSONRRNYVAVFALFVLGGTHAAAFHVQLHNGNLGGLLSGLQSLVMVSGFIGLI
orf47ng-1    MALPVIRSONRRNYVAVFAIFVLGGTHAAAFHVQLHNGNLGGLLSGLQSLVMVSGFIGLI
      130     140     150     160     170     180

      190     200     210     220     230     240
orf47-1.pep GTRIISFSTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMAGVLAWSAVFAFAAGVIFT
orf47ng-1    GMRIISFSTSKRLNVPQIPSPKWVAQASLWLPMLTAILMAHGVMPWLSAAFAFAAGVIFT
      190     200     210     220     230     240

      250     260     270     280     290     300
orf47-1.pep VQVYRWYKYPVLKEPMLWILFAGYLFTGLGLIAVGASYFKPAFLNLGVHLIGVGIGVLT
orf47ng-1    VQVYRWYKYPVLKEPMLWILFAGYLFTGLGLIAVGASYFKPAFLNLGVHLIGVGIGVLT
      250     260     270     280     290     300

      310     320     330     340     350     360
orf47-1.pep LGMMARTALGHTGNPIYPPPKAVPVAFWLMAATAVRMVAVFSSGTAYTHSIRTSSVLEA
orf47ng-1    LGMMARTALGHTGNSIYPPPKAVPVAFWLMAATAVRMVAVFSSGTAYTHSIRTSSVLEA
      310     320     330     340     350     360

      370     380
orf47-1.pep LALLVYAWKYIPWLIRPRSDGRPGX
orf47ng-1    LALLVYAWKYIPWLIRPRSDGRPGX
      370     380

```

Furthermore, ORF47ng-1 shows significant homology to an ORF from *Pseudomonas stutzeri*:

65 gnl|PID|e246540 (Z73914) ORF396 protein [*Pseudomonas stutzeri*] Length = 396  
Score = 155 bits (389), Expect = 5e-37

-408-

Identities = 121/391 (30%), Positives = 169/391 (42%), Gaps = 21/391 (5%)

Query: 7 PVWAMAFRPFYSLAALYGALSULLWGFYGTGTHELSGFY-----WHAHEMIWGYAGLV 59  
 P+W +AFRPF+ +LY L++ LW +TG GF WH HEM++G+A +  
 Sbjct: 14 PIWRLAFRPFFLAGSLYALLAIPLVAAWTGLWP--GFQPTGGWLAWHRHEMLFGFAMAI 71

Query: 60 VIAFLLTAVATWTGQPPTRGGLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAV 119  
 V FLLTAV TWTGQ G LVGL A WLAAR+ ++ G AA L LF  
 Sbjct: 72 VAGFLLTAVQTWTGQTAPSGNRLVGLAAVWLAARL-GWLFGLPAAWLAPLDLLFLVALVW 130

Query: 120 CMALPVIRSONRRNYVAVFAIFVLGGTHAAFXXXXXXXXXXXXXXXXXXXXXMVS GFGL 179  
 MA + + +RNY V + ++ G +V+ + L  
 Sbjct: 131 MMAQMLWAVRQKRNYPIVVVLSLMLGADVLIITGLLQGNDAQRQGVLAGLWLVAALMAL 190

Query: 180 IGMRIISFFTSKRLNVPQIPSP-KWVAQASLWLPMLTAILMAHGV----MPWLSAAFAFA 234  
 IG R+I FFT + L P W+ A L + A+L A GV P L F A  
 Sbjct: 191 IGGRVIPFFTQRLGKVDVAVKPPWVLDVALLVGTGVIALHAFGVAMRPQPLLGLLEFV-A 249

Query: 235 AGVIFTVQVYRWYKPVLEKPMWLIFAGYLFGLGLIAGVASYF-KPAFXXXXXXXXXXX 293  
 GV +++ RW+ K + K +LW L L+ + + +F A  
 Sbjct: 250 IGVGHLLRLMRWYDKGIWKVGLLWSLHVAMLWLVAAFGLALWHFGLLAQSSPSLHALSV 309

Query: 294 XXXXXXXXXXXMMARTALGHTGNSIYPPEKAVPVAFWLXXXXXXXXXXXXFSSGTAYTHSIR 353  
 M+AR LGHTG + P + AF L F S +  
 Sbjct: 310 GSMSGLILAMIAVRTLTGHTGRPLQPLAGIIG-AFVL---FNLGTAARVFLSVAPVGGGLW 365

Query: 354 TSSVLFALALLVYAWKYIPWLIRPRSDGRPG 384  
 ++V + LA +Y W+Y P L+ R DG PG  
 Sbjct: 366 LAAVCWTLAFALYVWRYAPMLVAARVDGHPG 396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 85

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 715>:

35 1 ..ATGCCGTCTG AAGGTTCAAG CGGCmTCGGT GyCGGGGAay CAGAAGyGGT  
 51 AGCGCATGCC CAATGAGACT TCGTGGGTTT TGAAGCGGGT GTTTTCCAAG  
 101 CGTCCCCAGT TGTGGTAACG GTATCCGGTG TCyAArGTCA GCTTGGGyGT  
 151 GATGTCGAAa CCGACACCGG CGATGACACC AAGACCyAmG CTGCTGATrC  
 201 TGTkGCTTTC GTGATAGGSa GGTtTGyTGG kmksAsyTTG TAYrATwkkG  
 40 251 CCTssCwstG kAGmGCCkTk CkyTGgTkkA swGrwArTAG TCGTGGTtTy  
 301 TkTtTyCACc GAATGAACyT GATGTTTAAC GTGTCCGTAG GCGACGCGCG  
 351 CGCCGATATA GGGTTTGAAT TTATCGTTGA GTTTGAAATC GTAAATGGCG  
 401 GACAAGCCGA GAGAAGAAAC GGCGTGGAAG CTGCCGTTTC CCTGATGTTT  
 451 TGTTTGGGTT TCTTTGTAGT TGTGTTTAT CTCTTCAGTA ACTTTTTTAG  
 45 501 TAGAAGAATT ACTTTCTTTC CATTTTCTGT AACTGGCATA ATCTGCCGCT  
 551 ATTCTCCAGC CGCCGAAATC ..

This corresponds to the amino acid sequence <SEQ ID 716; ORF67>:

50 1 ..MPSEGSdGXG XGEXEXVAHA QXDFVGFEG VEQASPVVVT VSGVXXQLGX  
 51 DVETDTGDDT KTXAADXVAF VIGRFxGXxL YXXAXXXxAX XWXXXxSRGF  
 101 XXHRMNLmFN VSVGDARADI GFEFIVEFEI VNGGQAERRN GVEAAVSLMF  
 151 CLGFEVvVvY LFSNFFSRRI TFFPFSVTGI ICRYSPAAEI ..

Computer analysis of this amino acid sequence gave the following results:

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF67 shows 51.8% identity over 199 aa overlap with a predicted ORF (ORF67ng) from

55 *N.gonorrhoeae*:

	orf67.pep		MPSEGS	SDGXGXGEXE	XVAHAQXDFVGF	EAG	30
	orf67ng	TNFEIAVL	SGMTVR	VFYCAR	PAPVNG	GRLKMPSEGS	146
5		90	100	110	120	130	140
	orf67.pep	VFQASPV	VTVSGV	XXQLGX	DVETDT	GDGDTKTX	90
	orf67ng	VFQASPV	VAVAGV	QGQAGR	DVYAHAR	HRAEAQAAA	206
10							
	orf67.pep	XWXXXSR	GFXH	RNM	MFNV	SVGDAR	150
	orf67ng	TRVGGK	STCYFF	SRIDAV	SDVSVG	DARTDIG	266
15							
	orf67.pep	CLGFFV	V-----	VVYLF	SNFFS	RRITFF-	190
	orf67ng	RLLVFY	VKLVA	AKSFII	LSFQ	LFYVHG	326

The ORF67ng nucleotide sequence <SEQ ID 717> is predicted to encode a protein comprising amino acid sequence <SEQ ID 718>:

20	1	MPSETVGSIV	NVGVD	ESVGF	SPPFPSIQHF	YRFHRIHRIR	LFRPPGPMQL
	51	NRHSHGSGNL	GRGVAT	VLVS	DKFPCGQVRI	PACAGMTNFE	IAVLSGMTVR
	101	VFYCARPAPV	NGGRLK	MPSE	GSDGIGIGES	EAVAHAQRGF	VGFEAGVFQA
	151	SPVVAVAGV	QGQAGR	DVYA	HARHRAEAQA	AAVAFLIGV	FLRMSVRINR
	201	NCCVSITRVG	GKSTCY	FFSR	IDAVSDVSVG	DARTDIGFEF	VVEFEIVNGG
	251	QAERRNGVEC	AVFLMF	RLLV	FYVKLVAAKS	FIILSFQLFY	VHGIFIVVPF
25	301	PVTGIIRGDA	FAAEV	VADRH	PGVDGMRTDV	SEITAYRAYF	VFAWSGWFR
	351	IVGNAFGG	VG	*			

Based on the presence of a several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### 30 Example 86

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 719>

	1	ATGTTTGCTT	TTTGA	AAGC	CTTTT	TGTC	GAATAC	GGTT	ATGCGG	CTGT
	51	TTTTTTTGTA	TTGGT	CATCT	GCGGT	TTCGG	CGTGCC	GATT	CCCGAG	GATT
35	101	TGACCTTGGT	AACAGG	CGGC	GTGATT	TTCGG	GTATGG	GTTA	TACCAAT	CCG
	151	CATATTATGT	TTGCAG	TCGG	TATGCT	CGGC	GTATTG	GTCG	GGAC	GGCAT
	201	CATGTTTCGCC	GCCGGA	CAGAA	TTTGGG	GGCA	GAAAT	CCCTA	AGGTT	CAAC
	251	CTATTGCGsG	CATCAT	GACG	CCGAA	ACGTT	ATGAGC	CAGGT	TCAGGA	AAAA
	301	TTCGACAAAT	ACGGTA	ACTG	GGTCTT	TATT	GTGCCC	CGTT	TCCTG	CCCG
	351	TTTGAGAACG	GCCGTAT	TTG	TTACAG	CCCG	TATCAG	CCGC	AAGGTT	TCAT
40	401	ACTTGCGTTT	TATCATT	TATG	GATGG	ACTGG	CCGA...			

This corresponds to the amino acid sequence <SEQ ID 720; ORF78>:

	1	MFAFLEAFFV	EYGYAA	VFFV	LVICGF	GVPI	PEDLT	LVTTG	VISGM	GYTNP
	51	HIMFAVGMLG	VLVGD	GIMFA	AGRIW	GQXXL	XFXPI	AXIMT	PXRYE	QVQEK
	101	FDKYNWVLF	VARFL	PGLRT	AVFVT	AGISR	KVSYL	RFIIM	DGLAA...	

45 Further work revealed the complete nucleotide sequence <SEQ ID 721>:

	1	ATGTTTGCTT	TTTGA	AAGC	CTTTT	TGTC	GAATAC	GGTT	ATGCGG	CTGT
	51	TTTTTTTGTA	TTGGT	CATCT	GCGGT	TTCGG	CGTGCC	GATT	CCCGAG	GATT
	101	TGACCTTGGT	AACAGG	CGGC	GTGATT	TTCGG	GTATGG	GTTA	TACCAAT	CCG
	151	CATATTATGT	TTGCAG	TCGG	TATGCT	CGGC	GTATTG	GTCG	GGAC	GGCAT
50	201	CATGTTTCGCC	GCCGGA	CAGAA	TTTGGG	GGCA	GAAAT	CCCTA	AGGTT	CAAC
	251	CTATTGCGCG	CATCAT	GACG	CCGAA	ACGTT	ATGAGC	CAGGT	TCAGGA	AAAA
	301	TTCGACAAAT	ACGGTA	ACTG	GGTCTT	TATT	GTGCCC	CGTT	TCCTG	CCCG
	351	TTTGAGAACG	GCCGTAT	TTG	TTACAG	CCCG	TATCAG	CCGC	AAGGTT	TCAT
	401	ACTTGCGTTT	TATCATT	TATG	GATGG	ACTGG	CCGA	ACTGAT	TTCCG	TCCCT
55	451	ATTGGA	TTT	ATCTG	GGCGA	ATACG	GTGCG	CACAAC	ATCG	ATTGCTGAT

-410-

501 GGCAGAAATG CACAGCCTGC AATCGGGTAT TTTTGTATC TTGGGTATAG  
 551 GTGCGACCGT TGTCGCTTGG ATTTGGTGA AAAACGCCA ACGTATCCAG  
 601 TTTTACCGCA GCAAATTGAA AGAAAAGCGG GCGCAACGCA AAGCCGCCAA  
 651 GGCAGCCAAA AAAGCCGCGC AAAGCAAACA ATAA

5 This corresponds to the amino acid sequence <SEQ ID 722; ORF78-1>:

1 MFAFLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP  
 51 HIMFAVGMLG VLVGDGIMFA AGRIWGQKIL RFKPIARIMT PKRYEQVQEK  
 101 FDKYGNWVLF VARFLPGLRT AVEVTAGISR KVSYLRFIIM DGLAALISVP  
 151 IWIYLGEYGA HNIDWLMAM HSLQSGIFVI LGIGATVVAW IWWKKRQRIQ  
 201 FYRSKLKEKR AQRKAACA KAAQSKQ\*

Computer analysis of this amino acid sequence predicts several transmembrane domains, and also gave the following results:

Homology with the dedA homologue of *H. influenzae* (accession number P45280)

ORF78 and the dedA homologue show 58% aa identity in 144aa overlap:

15 Orf78: 4 FLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGM--GYTNPHIMFAVGMLGV 61  
 FL FF EYGY AV FVL+ICGFGVPIPED+TLV+GGVI+G+ N H+M V M+GV  
 DedA: 20 FLIGFFTEYGYWAVLFLIICGFGVPIPEDITLVSGGVIAGLYPENVNHLMLLVSMIGV 79  
 20 Orf78: 62 LVGDGIMFAAGRIWGQXXLFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRTA 121  
 L GD M+ GRI+G L F PI I+T R V+EKE +YGN VLFVARFLPGLR  
 DedA: 80 LAGDSCMYWLGRIYGTILRFRPIRIRIVTLQRLRMVREKFSQYGNRVLFVARFLPGLRAP 139  
 Orf78: 122 VEVTAGISRKVSYLRFIIMDGLAA 145  
 +++ +GI+R+VSY+RF+++D AA  
 25 DedA: 140 IYMVSGITRRVSYVRFVLIDFCAA 163

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF78 shows 93.8% identity over a 145aa overlap with an ORF (ORF78a) from strain A of *N. meningitidis*:

30 orf78.pep 10 20 30 40 50 60  
 MFALFLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG  
 orf78a MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG  
 35 10 20 30 40 50 60  
 orf78.pep 70 80 90 100 110 120  
 VLVGDGIMFAAGRIWGQXXLFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRT  
 orf78a VLVGDGIMFAAGRIWGQKILFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT  
 40 70 80 90 100 110 120  
 orf78.pep 130 140  
 AVEVTAGISRKVSYLRFIIMDGLAA  
 45 orf78a AVEVTAGISRKVSYLRFIIMDGLAALISVPVWIYLGEYGAHNIDWLMAMHSLQSGIFIA  
 130 140 150 160 170 180

The complete length ORF78a nucleotide sequence <SEQ ID 723> is:

1 ATGTTTGCCC TTTTGAAGC CTTTTTTGTC GAATACGGCT ATGCGGCCGT  
 51 GTTTTTCGTT TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCGGAGGATT  
 101 TGACCTTGGT AACAGGCGGC GTGATTTCGG GTATGGGTTA TACCAATCCG  
 151 CATATTATGT TTGCAGTCGG TATGCTCGGC GTATTGGTCG GGGACGGCAT  
 201 CATGTTCCGC GCCGGACGCA TCTGGGGGCA GAAATCCTC AAGTTCAAAC  
 251 CGATTGCGCG CATCATGACG CCGAAACGTT ACGCACAGGT TCAGGAAAAA  
 301 TTCGACAAAT ACGGCAACTG GGTGTTATTT GTCGCTCGTT TCCTGCCCGG  
 351 TTTGCGGACT GCGGTTTTCG TTACCGCCGG CATCAGCCGC AAAGTATCGT  
 401 ATCTGCGCTT TCTGATTATG GACGGGCTTG CCGCGCTGAT TTCCGTGCCC  
 451 GTTTGGATT ACTTGGGCGA GTACGGCGCG CACAACATCG ATTGGCTGAT

-411-

501 GGCAGAAATG CACAGCCTGC AATCCGGCAT CTTCATCGCA TTGGGCGTGC  
 551 TGGCGGCGGC GCTGGCGTGG TTCTGGTGGC GCAAACGCCG ACATTATCAG  
 601 CTTTACCGCG CACAATTGAG CGAAAAACGC GCCAAACGCA AGGCGGAAAA  
 651 GGCAGCGAAA AAAGCGGCAC AGAAGCAGCA GTAA

5 This encodes a protein having amino acid sequence <SEQ ID 724>:

1 MFALLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP  
 51 HIMFAVGMLG VLVDGIMFA AGRIWGQKIL KFKPIARIMT PKRYAQVQEK  
 101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRLIM DGLAALISVP  
 151 VWIYLGEYGA HNIDWLMAM HSLQSGIFIA LGVLAALAW FWRKRHHYQ  
 201 LYRAQLSEKR AKRKAERKAA KAAQKQQ\*

ORF78a and ORF78-1 show 89.0% identity in 227 aa overlap:

		10	20	30	40	50	60
15	orf78a.pep	MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG					
	orf78-1	MFAFLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG					
		10	20	30	40	50	60
20	orf78a.pep	VLVDGIMFAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT	70	80	90	100	110
	orf78-1	VLVDGIMFAAGRIWGQKILRFKPIARIMTPKRYEQVQEKFDKYGNWVLFVARFLPGLRT	70	80	90	100	110
		70	80	90	100	110	120
25	orf78a.pep	AVFVTAGISRKVSYLRLIMDGLAALISVPVWVYLGEYGAHNIDWLMAMHSLQSGIFIA	130	140	150	160	170
	orf78-1	AVFVTAGISRKVSYLRFIIMDGLAALISVPVWVYLGEYGAHNIDWLMAMHSLQSGIFVI	130	140	150	160	170
		130	140	150	160	170	180
30	orf78a.pep	LGVLAALAWFWRKRHHYQLYRAQLSEKRAKRAKAAKAAQKQQX	190	200	210	220	
	orf78-1	LGIGATVVAWIWKKRQRIQFYRSKLKEKRAQRKAAKAAKAAQSKQX	190	200	210	220	
35		190	200	210	220		

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF78 shows 97.4% identity over 38 aa overlap with a predicted ORF (ORF78ng) from *N. gonorrhoeae*:

40	orf78.pep	XXLXFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRTAVFVTAGISRKVSYLRF	137
	orf78ng	YPVLFVARFLPGLRTAVFVTAGISRKVSYLRF	32
	orf78.pep	IIMDGLAA	145
45	orf78ng	LIMDGLAALISVPVWVYLGEYGAHNIDWLMAMHSLQSGIFIALGVLAALAWFWRKR	92

The ORF78ng nucleotide sequence <SEQ ID 725> is predicted to encode a protein comprising amino acid sequence <SEQ ID 726>:

1 ..YPVLFVARFL PGLRTAVFVT AGISRKVSYL RFLIMDGLAA LISVPVWVYL  
 51 GEYGAHNIDW LMAKMHSLS GIFIALGVLA AALAWFWRK RRHYQLYRAQ  
 101 LSEKRAKRA EKAAKKAAQK QQ\*

Further work revealed the complete gonococcal nucleotide sequence <SEQ ID 727>:

1 atgttttgcct tttTggaagc CTTTTTTGTC GAAtacggCt atgcGGCCGT  
 51 GTTTTTTCGT TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCCGAAGATT  
 101 TGACCTTGGT AACGGGCGGC GTGATTTCGG GTATGGGTTA TACCAATCCG  
 151 CATATTATGT TTGCGGTCGG TATGCTCGGC GTGTTGGCGG GCGACGGCGT  
 201 GATGTTTGCC GCCGGACGCA TCTGGGGGCA GAAAATCCTC AAGTTCAAAC  
 251 CGATTGCGCG CATCATGACG CCGAAACGTT ACGCGCAGGT TCAGGAAAAA  
 301 TTCGACAAAT ACGGCAACTG GGTCTGTCTT GTCGCCCGTT TCCTGCCGGG



This corresponds to the amino acid sequence <SEQ ID 728; ORF78ng-1>:

1	MFALLEAFFV	EYGYAAVFFV	LVICGFGVPI	PEDLTLVTGG	VISGMGYTNP
51	HIMFAVGMGLG	VLADGVMFA	AGRWGQKIL	KFKPIARIMT	PKRYAQVQEK
101	FDKYGNWVLF	VARFLPLGLT	AVFTAGISIR	KVSYLRLFLM	DGLAALISVP
151	VWIIYLGEYGA	HNIDWLMAKM	HSLSQGFIA	LGVLAAALAW	FWWRKRHHYQ
201	LYRAQLSEKR	AKRKAEKAAC	KAAQKQ*		

15

20

25

30

35

40

45

50

55

60

5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 729>:

15 This corresponds to the amino acid sequence <SEQ ID 730; ORF79>:

Further work revealed the complete nucleotide sequence <SEQ ID 731>:

20 1 ATGAAAAAAT TATTGGCGGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT  
51 TTCCGCCGCC GGAGTCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG  
101 AAGGTATGAA AATAGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC  
151 AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCCGTTGCCG ACCGCGTCGA  
25 201 AGTGCATACC CACATCAACG ACAACGGCGT GCGTGGGATG CGCGAAGCTCG  
251 AAGGCGGCGT GCCTTTGGAA GCGAAATCCG TTACCGAACT CAAACCCGGC  
301 AGCTATCATG TGATGTTTAT GGGTTTGAAG AAACAATTAA AAGAGGGCGA  
351 TAAAAATCCC GTTACCCTGA AATTTAAAAA CGCCAAAGCG CAAACCGTCC  
401 AACTGGAAGT CAAAATCGCG CCGATGCCGG CAATGAACCA CGGTATCAC  
451 CACGGCGAAG CGCATCAGCA CTAA

30 This corresponds to the amino acid sequence <SEQ ID 732; ORF79-1>:

1 MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKIGG AFMKIHND EA  
51 KQDFLLGSS PVADRVEVHT HINDNGVMRM REVEGGVPLE AKSVTE LKPG  
101 SYHVMFMGLK QQLKEGD KIP VTLKFKNAKA QTVQLEV KIA PMPAMNHGHH  
151 HGEAHOH\*

35 Computer analysis of this amino acid sequence revealed a putative leader peptide and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF79 shows 94.6% identity over a 147aa overlap with an ORF (ORF79a) from strain A of *N. meningitidis*:

40		10	20	30	40	50	60
	orf79.pep	<u>MKKLLAAVMMAGLAGAVSAAGVHVEDGWARTTVEGMKIGGAFMKIHND EAKQDFLLGGSS</u>					
	orf79a	<u>MKXLLAAVMMAGLAGAVSAAGIHVEDGWARTTVEGMKMGGA FMKIHND EAKQDFLLGGSS</u>					
45		10	20	30	40	50	60
	orf79.pep	70	80	90	100	110	120
		PVADRV E VH THINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP					
	orf79a	PVADRV E VH THINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGXKKQLKXGDKIP					
50		70	80	90	100	110	120

-414-

```

              130      140
orf79.pep    VTLKFKNKAQTQVQLEVKIAPMPAMNH
5            |||||
orf79a       VTLKFKNKAQTQVQLEVKTAPMSAMDHGHHHGEAHQH
              130      140      150

```

The complete length ORF79a nucleotide sequence <SEQ ID 733> is:

```

1  ATGAAANAAC TATTGGCAGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
51 TTCCGCCGCC GGAATCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG
10  AAGGTATGAA AATGGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC
101 AACAAGACT TTTGCTCGG CGGAAGCAGC CCTGTTGCCG ACCGCGTCGA
151 AGTGCATACC CATATCAATG ATAACGGTGT GATGCGGATG CGCGAAGTCG
201 AAGGCGGCGT GCCTTTGGAG GCGAAATCCG TTACCGAACT CAAACCGGCG
251 AGCTATCATG TCATGTTTAT GGGTNTGAAA AAACAATTAA AAGANGGCGA
301 CAAGATCCC GTTACCCTGA AATTTAAAAA CGCAAAGCA CAAACCGTCC
15  401 AACTGGAAGT CAAACCGCG CCGATGTCGG CAATGGACCA CGGTCATCAC
451 CACGGCGAAG CGCATCAGCA CTAA

```

This encodes a protein having amino acid sequence <SEQ ID 734>:

```

1  MKXLLAAVMM AGLAGAVSAA GIHVEDGWAR TVEGMKMGG AFMKIHNDEA
20  51 KQDFLLGGSS PVADRVEVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
101 SYHVMFMGXK KQLKXGDKIP VTLKFKNKA QTQVQLEVKA PMSAMDHGHH
151 HGEAHQH*

```

ORF79a and ORF79-1 show 94.9% identity in 157 aa overlap:

```

25  orf79a.pep    10      20      30      40      50      60
      MKXLLAAVMMAGLAGAVSAAGIHVEDGWAR TVEGMKMGGA FMKIHNDEAKQDFLLGGSS
      |||||
orf79-1  MKXLLAAVMMAGLAGAVSAAGVHVEDGWAR TVEGMKIGGA FMKIHNDEAKQDFLLGGSS
              10      20      30      40      50      60
30  orf79a.pep    70      80      90      100     110     120
      PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGXKKQLKXGDKIP
      |||||
orf79-1  PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
              70      80      90      100     110     120
35  orf79a.pep    130     140     150
      VTLKFKNKAQTQVQLEVKTAPMSAMDHGHHHGEAHQH
      |||||
orf79-1  VTLKFKNKAQTQVQLEVKIAPMPAMNHGHHHGEAHQH
              130     140     150
40

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF79 shows 96.1% identity over 76 aa overlap with a predicted ORF (ORF79ng) from *N.gonorrhoeae*:

```

45  orf79.pep    FMKIHNDEAKQDFLLGGSSPVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGS 101
      |||||
orf79ng                                INDNGVMRMREVKGVPLEAKSVTELKPGS 30
50  orf79.pep    YHVMFMGLKKQLKEGDKIPVTLKFKNKAQTQVQLEVKIAPMPAMNH 147
      |||||
orf79ng    YHVMFMGLKKQLKEGDKIPVTLKFKNKAQTQVQLEVKTAPMSAMNHGHHHGEAHQH 86

```

An ORF79ng nucleotide sequence <SEQ ID 735> was predicted to encode a protein comprising amino acid sequence <SEQ ID 736>:

```

1  ..INDNGVMRM EVKGGVPLEA KSVTELKPGS YHVMFMGLKK QLKEGDKIPV
51 TLKFKNKAQ TVQLEVKTAP MSAMNHGHHH GEAHQH*

```

Further work revealed the complete gonococcal DNA sequence <SEQ ID 737>:

1 ATGAAAAAAT TATTGGCAGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT  
 51 TTccgccgCc GGagTccAtG TCGAggACGG CTGGGCGCGc accaCTGtcg  
 101 aaggtATgaa aatggGCGGC GCgttCATga aaATCCACAA CGACGaaGcc  
 151 atacaaGACt ttgtgcTCgg CGGaagcatg cccgttgccg accgcGTCGA  
 5 201 AGTGCAtaca cacATCAACG ACAACGGCGT GATGCGTATG CGCGAAGTCA  
 251 AAGGCGGCGT GCCTTTGGAG GCGAAATCCG TTACCGAACT CAAACCCGGC  
 301 AGCTATCACG TGATGTTTAT GGGTTTGAAA AAACAACGTA AAGAGGGCGA  
 351 CAAGATTCCC GTTACCCCTGA AATTTAAAAA CGCCAAAGCG CAAACCGTCC  
 10 401 AACTGGAAGT CAAAACCGCG CCGATGTCGG CAATGAACCA CGGTCATCAC  
 451 CACGGCGAAG CGCATCAGCA CTAA

This corresponds to the amino acid sequence <SEQ ID 738; ORF79ng-1>:

1 MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKMG GAFMKIHND EAFMKIHND EAFMKIHND  
 51 IQDFVLG GSM PVADRVEVHT HINDNGVM RM REVKG GVPLE AKSVTELK PG  
 101 SYHVMFM GLK KQLKEGDK IP VTLKFKNA KA QTVQLEV KTA PMSAMNH GH  
 15 151 HGEAHQH\*

ORF79ng-1 and ORF79-1 show 95.5% identity in 157 aa overlap:

		10	20	30	40	50	60
20	orf79-1.pep	MKKLLAAVMMAGLAGAVSAA	GVHVEDGWAR	TTVEGMKIGGAFMKIHND	EAFMKIHND	EAFMKIHND	EAFMKIHND
	orf79ng-1	MKKLLAAVMMAGLAGAVSAA	GVHVEDGWAR	TTVEGMKMGGA	AFMKIHND	EAFMKIHND	EAFMKIHND
		10	20	30	40	50	60
		70	80	90	100	110	120
25	orf79-1.pep	PVADRVEVHTHINDNGVMRM	REVEGGVPLEAKSVTELKPG	SYHVMFMGLKKQLKEGDKIP			
	orf79ng-1	PVADRVEVHTHINDNGVMRM	REVKGVPLEAKSVTELKPG	SYHVMFMGLKKQLKEGDKIP			
		70	80	90	100	110	120
30		130	140	150			
	orf79-1.pep	VTLKFKNAKAQTVQLEVKI	APMPAMNHGHHHGEAHQH				
	orf79ng-1	VTLKFKNAKAQTVQLEVKI	APMSAMNHGHHHGEAHQH				
		130	140	150			

Furthermore, ORF79ng-1 shows significant homology to a protein from *Aquifex aeolicus*:

35 gi|2983695 (AE000731) putative protein [Aquifex aeolicus] Length = 151  
 Score = 63.6 bits (152), Expect = 6e-10  
 Identities = 38/114 (33%), Positives = 58/114 (50%), Gaps = 1/114 (0%)  
 40 Query: 24 VEDGWAR TTVEGMKMGGA AFMKIHND EAFMKIHND EAFMKIHND EAFMKIHND  
 V+ W G M I N+ D+++G +A RVE+H + +N V +M  
 Sbjct: 27 VKHPVMEPPPPGNTTMMGMII VNEGDEPDY LIGAKTDIAQRVELHKT VTIENDVAKMVPQ 86  
 Query: 84 KGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEV 137  
 + + + K E K YHVM +GLKK++KEGDK+ V L F+ + TV+ V  
 45 Sbjct: 87 ER-IEIPKKGKVEFKHHGYHVMII GLKKRIKEGDKVKVELIFEKSGKITVEAPV 139

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

50 ORF79-1 (15.6kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 18A shows the results of affinity purification of the His-fusion protein. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 18B) These experiments confirm that ORF79-1 is a surface-exposed protein, and that it is a useful immunogen.

1	ATGACGGTAA	CTGCGGCCGA	AGGCGGCAAA	GCTGCCAAGG	CGTTAAAAAA
51	ATATCTGATT	ACGGGCATT	TGGTCTGGCT	GCCGATTGCG	GTAAACGGTTT
101	GGGTGGTTTC	CTATATCGTT	TCCGCGTCCG	ATCAGCTCGT	CAACCTGCTG
151	CCGAAGCAAT	GGCGGCCGCA	ATATGTTTTG	GGGTTTAATA	TCCCGGGGCT
201	GGGCGTTATC	GTTGCCATTG	CCGTATTGTT	TGTAACCGGA	TTGTTTGCCG
251	CCAACGTATT	GGGTCCGCAG	ATCCTCGCCG	CGTGGGACAG	CCTGTTGGGG
301	CGGATTCGG	TTGTGAAATC	CATCTATTCG	AGTGTGAAAA	AAGTATCCGA
351	ATacgTGCTG	TCCGACGACA	GCGGTTCTGT	TAAACGCGCG	GTACTCGTGC
401	CGTTTCCCA	GCCCGBTATT	TGGACGATyG	CTTTCGTGTC	AGGGCAGGTG
451	TGCAATGCG	TTAAGGCCGC	ATTGCCGAAS	GACGGCGATT	ATCTTTCCGT
501	GCTATGTTCCG	ACCACGCCGA	ATCCGACCGG	CGGTACTACT	ATTATGGTAA
551	AGAAAAGCGA	TGTGCGCGAA	CTCGATATGA	GCGTGGACGA	AsCATTGAAA
601	TATGTGATTT	CGCTGGGTAT	GGTCATCCCT	GACGACCTGC	CCGTCAAAC
651	ATTGGCAsGA	CCTATGCCGT	CTGAAAAGGC	GGATTGCCCC	GAACAACAAT
701	AA				

```

20      1  MTVTAAEGGK  AAKALKKYLI  TGI L VWLPIA  VTVVVS YIV  SASDQLVNLL
      51  PKQWRPQYVL  GFNIPGLGVI  VAI AVLFTG  LFAANVLGRQ  ILAAWDSLLG
     101  RIPVVKSIYS  SVKKVSEYVL  SDSRSFKTP  VLVFPQPQGI  WTIAFVSGQV
     151  SNAVKAAALPX  DDYLSVYVP  TTPNPTGGYY  IMVKKSDVRE  LDMSVDExLK
     201  YVISLGMVIP  DDL PVKTLAX  PMPSEKADLP  EQO*

```

	1	ATGACGGAAC	nTGGCGGCCGA	AGGCGGCAAA	GCTGCCAARg	CGTTAAAAAA
	51	ATATCTGATT	ACGGGCATTT	TGGTCTGGCT	GCCGATTGCG	GTACGGTTT
	101	GGTGGTTTC	CTATATCGTT	TCCGCGTCCG	ATCAGCTCGT	CAACCTGCTG
30	151	CCGAAGCAAT	GGCGGCCGCA	ATATGTTTTG	GGGTTTAATA	TCCCGGGGCT
	201	GGGCGTTATC	GTTGCCATTG	CCGTATTGTT	TGTAACCGGA	TTGTTTGCCG
	251	CCAACGTATT	GGGTCCGGCAG	ATCCTCGCCG	CGTGGGACAG	CCTGTTGGGG
	301	CGGATTCCGG	TTGTGAAATC	CATCTATTCC	AGTGTGAAAA	AAGTATCCGA
	351	ATCGCTGCTG	TCCGACAGCA	GCCGTTCGTT	TAAACGCCCG	GTACTCGTGC
	401	CGTTTCCCCA	GCCCCGTATT	TGGACGATTG	CTTTCGTGTC	AGGGCAGGTG
35	451	TCGAATCGCG	TTAAGGCCGC	ATTGCCGAAG	GACGGCGATT	ATCTTTCCGT
	501	GTATGTTCCG	ACCAGCCGCA	ATCCGACCCG	CGGTTACTAT	ATTATGGTAA
	551	AGAAAAGCGA	TGTGCGCGAA	CTCGATATGA	GCGTTGACGA	AGCATTGAAA
	601	TATGTGATTT	CGCTGGGTAT	GGTCATCCCT	GACGACCTGC	CCGTCAAAC
	651	ATTGGCAGGA	CCTATGCCGT	CTGAAAAGGC	GGATTGCCC	GAACAACAAT
40	701	AA				

45

1	MTEXAAEGGK	AAKALKKYLI	TGILVWLPIA	VTWVVVSIVY	SASDQLVNLL
51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFTVG	LFAANVLGRQ	ILAAWDSLLG
101	RIPVVKSIYS	SVKKVSESL	SDSSRSFKTP	VLPFPQPGI	WTIAFVSGQV
151	SNAAKAAALPK	DGKYLVSVPY	TTNPNTGGYY	IMVKKSDVRE	LDMSVDEALK
201	YVISLGMVIP	DDL PVKTLAG	PMPSEKADLP	EQO*	

### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF98 shows 96.1% identity over a 233aa overlap with an ORF (ORF98a) from strain A of *N.*

```

              10      20      30      40      50      60
orf98.pep    MVTAAEGGKAAKALKKYLITGILVWLPIAVTVVWVSYIVSASDQLVNLLPKQWRPQYVL
              ||  |||||
orf98a       MTEPAAEGGKAAKALKKYLITGILVWLPIAVTVVWVSYIVSASDQLVNLLPKQWRPQYVL
              10      20      30      40      50      60

```

-417-

		70	80	90	100	110	120
	orf98.pep	GFNIPGLGVIVAI	AVLFVTGLFAANVLGRQILA	AWDSLGRIPVVKSIYSSVKKVSEYVL			
5	orf98a	GFNIPGLGVIVAI	AVLFVTGLFAANVLGRQILA	AWDSLGRIPVVKSIYSSVKKVXSLL			
		70	80	90	100	110	120
		130	140	150	160	170	180
10	orf98.pep	SDSSRSFKTPVLV	PFPPQGIWTIAFVSGQV	SNVKAALPKDGDYLSVYVPTT	PNPTGGYY		
	orf98a	SDSSRSFKTPVLV	PFPPQGIWTIAFVSGQV	SNVKAALPKDGDYLSVYVPTT	PNPTGGYY		
		130	140	150	160	170	180
15		190	200	210	220	230	
	orf98.pep	IMVKKSDVRELDMSV	DEALKYVISLGMVIPDDL	PVKTLAXPMPSEKADL	PEQQX		
	orf98a	IMVKKSDVRELDMSV	DEALKYVISLGMVIPDDL	PVKTLAGPMPSEKADL	PEQQX		
		190	200	210	220	230	

The complete length ORF98a nucleotide sequence <SEQ ID 743> is:

20	1	ATGACGGAAC	CTGCGGCCGA	AGGCGGCAAA	GCTGCCAAGG	CGTTAAAAAA
	51	ATATCTGATT	ACGGGCATTT	TGGTCTGGCT	GCCGATTGCG	GTAACGGTTT
	101	GGGTGGTTTC	CTATATCGTT	TCCGCGTCCG	ATCAGCTCGT	CAACCTGCTG
	151	CCGAAGCAAT	GGCGGCCGCA	ATATGTTTTG	GGGTTTAATA	TCCCGGGGCT
25	201	GGGCGTTATC	GTTGCCATTG	CCGTATTGTT	TGTAACCGGA	TTATTGCCG
	251	CAAACGTATT	GGCGCCGAG	ATTCTTGCCG	CGTGGGACAG	CTTGTTGGGG
	301	CGGATTCCGG	TTGTGAAGTC	CATCTATTCG	AGTGTGAAAA	AAGTATCCGA
	351	NTCGTTGCTG	TCCGACAGCA	GCCGTTCGTT	TAAAACACCA	GTACTCGTGC
	401	CGTTTCCCA	ATCGGTATT	TGGACAATCG	CATTCTGTGC	CGGTCAGGTG
	451	TCCAATGCGG	TTAAGCCGC	ATTGCCGAAG	GACGGCGATT	ATCTTCCGT
30	501	GTATGTTCCG	ACCACGCCGA	ATCCGACCGG	CGGTTACTAT	ATTATGGTAA
	551	AGAAAAGCGA	TGTGCGCGAA	CTCGATATGA	GCGTGGACGA	AGCGTTGAAA
	601	TATGTGATTT	CGCTGGGTAT	GGTCATCCCT	GACGACCTGC	CCGTCAAAC
	651	ATTGCGAGGA	CCTATGCCGT	CTGAAAAGGC	GGATTTGCCC	GAACAACAAT
	701	AA				

35 This encodes a protein having amino acid sequence <SEQ ID 744>:

	1	MTEPAAEGGK	AAKALKKYLI	TGILVWLPIA	VTWVVSYIV	SASDQLVNLL
	51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLEVTG	LFAANVLGRQ	ILAAWDSLIG
	101	RIPVVKSIYS	SVKKVXSLL	SDSSRSFKTP	VLVFPQSGI	WTIAFVSGQV
	151	SNVKAALPK	DGDYLSVYVP	TTPNPTGGYY	IMVKKSDVRE	LDMSVDEALK
40	201	YVISLGMVIP	DDLVPKTLAG	PMPSEKADLP	EQQ*	

ORF98a and ORF98-1 show 98.7% identity in 233 aa overlap:

		10	20	30	40	50	60
	orf98a.pep	MTEPAAEGGKA	AKALKKYLI	TGILVWLPIA	VTWVVSYIV	SASDQLVNLL	PKQWRPQYVL
45	orf98-1	MTEPAAEGGKA	AKALKKYLI	TGILVWLPIA	VTWVVSYIV	SASDQLVNLL	PKQWRPQYVL
		10	20	30	40	50	60
		70	80	90	100	110	120
50	orf98a.pep	GFNIPGLGVIVAI	AVLFVTGLFAANVLGRQILA	AWDSLGRIPVVKSIYSSVKKVXSLL			
	orf98-1	GFNIPGLGVIVAI	AVLFVTGLFAANVLGRQILA	AWDSLGRIPVVKSIYSSVKKVSESLL			
		70	80	90	100	110	120
		130	140	150	160	170	180
55	orf98a.pep	SDSSRSFKTPVLV	PFPPQGIWTIAFVSGQV	SNVKAALPKDGDYLSVYVPTT	PNPTGGYY		
	orf98-1	SDSSRSFKTPVLV	PFPPQGIWTIAFVSGQV	SNVKAALPKDGDYLSVYVPTT	PNPTGGYY		
		130	140	150	160	170	180
60		190	200	210	220	230	
	orf98a.pep	IMVKKSDVRELDMSV	DEALKYVISLGMVIPDDL	PVKTLAGPMPSEKADL	PEQQX		
	orf98-1	IMVKKSDVRELDMSV	DEALKYVISLGMVIPDDL	PVKTLAGPMPSEKADL	PEQQX		
		190	200	210	220	230	
65							

Homology with a predicted ORF from *N.gonorrhoeae*

ORF98 shows 95.3% identity over a 233 aa overlap with a predicted ORF (ORF98ng) from *N.gonorrhoeae*:

```

5      orf98.pep      10      20      30      40      50      60      60
      orf98ng      MTVPAAEGGKAALKKKYLITGILVWLPPIAVTVVWVSIVSASDQLVNLLPKQWRPQYVL      60
      orf98ng      MTEPAAEGGKAALKKKYLITGILVWLPPIAVTVVWVSIVSASDQLVNLLPKQWRPQYVL      60

10     orf98.pep      GFNIPGLGVIVAI AVLFTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSEYVL      120
      orf98ng      GFNIPGLGVIVAI AVLFTGLFAANVLGRQILAAWDSLLXRI PVVKSIYSSVKKVSESL      120

      orf98.pep      SDSSRSFKTPVLVFPFQPGIWTIAFVSGQV SNAVKAALPXDGDYLSVYVPTTNPNTGGYY      180
      orf98ng      SDSSRSFKTPVLVFPFQSGIWTIAFVSGQV SNAVKAALPQDGDYLSVYVPTTNPNTGGYY      180

15     orf98.pep      IMVKKSDVRELDMSVDEALKYVISLGMVIPDDL PVKTLAXPMPSEKADLPEQQ      233
      orf98ng      IMVKKSDVRELDMSVDEALKYVISLGMVIPDDL PVKTLAGPMPPEKAELPEQQ      233

```

20 The complete length ORF98ng nucleotide sequence <SEQ ID 745> is predicted to encode a protein having amino acid sequence <SEQ ID 746>:

```

1      MTEPAAEGGK AAKALKKKYL TGILVWLPPIA VTVWVVSIV SASDQLVNLL
51     PKQWRPQYVL GFNIPGLGVI VAI AVLFTG LFAANVLGRQ ILAAWDSLLX
25     101     RIPVVKSIYS SVKKVSESL SDSSRSFKTP VLVFPFQSGI WTIAFVSGQV
      151     SNAVKAALPQ DGDYLSVYVP TTPNPTGGYY IMVKKSDVRE LDMSVDEALK
      201     YVISLGMVIP DDL PVKTLAG PMPPEKAELP EQQ*

```

Further work revealed the complete nucleotide sequence <SEQ ID 747>:

```

1      ATGACGGAAC CTGCGGCCGA AGCGGCCAAA GCTGCCAAGG CGTAAAAAAA
30     51     ATATCTGATT ACAGGCATTT TGGTCTGGCT GCCGATTGCG GTAACGGTTT
      101     GGGTGGTTTC CTATATCGTT TCCGCGTCCG ACCAGCTTGT CAACCTGCTG
      151     CCGAAGCAAT GCGCGCCGCA ATATGTTTTG GGGTTTAATA TCCCCGGGCT
      201     CCGCGTTATT GTTGCCATTG CCGTATTGTT TGTAACCGGA TTATTTGCCG
      251     CAAACGTGTT GGGCCGGCAG ATTCTTGCCG CGTGGGACAG CCTGTTgggg
35     301     cggattCCGG TTGTCAAATC CATCTATTCG AGTGTGAAAA AAGTATCCGA
      351     ATCGCTGCTG TCCGACAGCA GCCGTTTCGT TAAACGCCG GTACTCGTGC
      401     CGTTTCCCA ATCGGGTATT TGGACAATCG CATTCGTGTC CGGTCAGGTG
      451     TCGAATGCGG TTAAGGCCGC ATTGCCGCGG GATGGCGATT ATCTTTCCGT
      501     GTATGTCCCG ACCACGCCCA ACCCGACCGG CGGTTACTAT ATTATGGTAA
40     551     AGAAAAGCGA TGTGCGCGAA CTCGATATGA GCGTGGACGA AGCGTTGAAA
      601     TATGTGATTT CGCTGGGTAT GGTCAATCCCT GACGACCTGC CCGTCAAAAC
      651     ATTGGCAGGA CCTATGCCGC CTGAAAAGGC GGAGTTGCCC GAACAACAAT
      701     AA

```

This corresponds to the amino acid sequence <SEQ ID 748; ORF98ng-1>:

```

45     1      MTEPAAEGGK AAKALKKKYL TGILVWLPPIA VTVWVVSIV SASDQLVNLL
      51     PKQWRPQYVL GFNIPGLGVI VAI AVLFTG LFAANVLGRQ ILAAWDSLLG
      101     RIPVVKSIYS SVKKVSESL SDSSRSFKTP VLVFPFQSGI WTIAFVSGQV
      151     SNAVKAALPQ DGDYLSVYVP TTPNPTGGYY IMVKKSDVRE LDMSVDEALK
      201     YVISLGMVIP DDL PVKTLAG PMPPEKAELP EQQ*

```

ORF98ng-1 and ORF98-1 show 97.9% identity in 233 aa overlap:

```

50     orf98-1.pep      10      20      30      40      50      60
      orf98ng-1      MTEPAAEGGKAALKKKYLITGILVWLPPIAVTVVWVSIVSASDQLVNLLPKQWRPQYVL
      orf98ng-1      MTEPAAEGGKAALKKKYLITGILVWLPPIAVTVVWVSIVSASDQLVNLLPKQWRPQYVL

55     orf98-1.pep      70      80      90      100     110     120
      orf98-1.pep      GFNIPGLGVIVAI AVLFTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSESL

```

-419-

5	orf98ng-1	GFNIPGLGVIVAI	AVLFVTGLFAAN	VLGRQILAAWDS	LLGRIPVVKSIY	SSSVKKVSESL	
		70	80	90	100	110	120
10	orf98-1.pep	SDSSRSFKTPVL	VFPFQPGIWTIA	FSVSGQVSNVKA	ALPKDGDYLSV	YVPTTPNPTGG	YY
	orf98ng-1	SDSSRSFKTPVL	VFPFQSGIWTIA	FSVSGQVSNVKA	ALPKDGDYLSV	YVPTTPNPTGG	YY
15		130	140	150	160	170	180
		190	200	210	220	230	
20	orf98-1.pep	IMVKKSDVRELD	MSVDEALKYVIS	LGMVIPDDL	LPVKTLAGPMP	SEKADLPEQQX	
	orf98ng-1	IMVKKSDVRELD	MSVDEALKYVIS	LGMVIPDDL	LPVKTLAGPMP	PEKAELPEQQX	

- Based on this analysis, including the fact that the putative transmembrane domains in the gonococcal protein are identical to the sequences in the meningococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 89

- The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 749>:

1	ATgAAAACGG	TAGTCTGGAT	TGTCGTCCTG	TTTGCCGCCG	CCGTCGGACT
51	GGCGCTGGCT	TCGGGCATTT	ACACCGGCGA	CGTGTATATC	GTA CTGGAC
101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTCGCTGATT
151	GCCGTCGTGG	TGTGGTATTT	CTTGTTTAAA	TTCATTATCG	G <sub>5</sub> GgTACTCA
201	ATATCCCCGA	AAAGATGCAG	CGTTTCGGTT	CGGCnCGTAA	AGGCCkCAAG
251	ssCGsGCTTG	CCTTGAACAA	GGCGGGTTTG	GCGTATTTTG	AAGGGCGTTT
301	TGAAAAGGCG	GAAGTAGAAG	CCTCACGCGT	GTTGGTCAAC	AAAGtAGGCC
351	GaGAGACAAC	CGGACTTTGG	CATTGATGCT	GrGCGCGCAC	GCCGCCGAC
401	AGATGGAAAA	CATCGAsTG	CGCGACCGTT	ATCTTGCGGA	AATCGCCAAA
451	CTGCCGGA	AAACAGCAGCT	TTCCCGTTAT	CTTTTGTTGG	CGGAATCGGC
501	GTTGAACCGG	CGCGATTACG	AAGCGGCGGA	AGCCAATCTT	CATGCGGCGG
551	CGAAGATGAA	TGCCAACCTT	ACGCGCCTCG	TGCGTCTGCA	.ATTCTGTAC
601	GCTTTCGACA	GGGGCGACGC	GTTGCAGGTT	CTGGCAAAAA	CCGAAAACT
651	TTCCAAGGCG	GGCGCGTTGG	GCAAATCGGA	AATGGAACGG	TATCAAAATT
701	GGGCATATCC	GTCGCCAGCT	GGCGGATGCT	GCCGATGCCG	CCGCTTTGAA
751	AACCTGCCTG	AAGCGGATTTC	CCGACAGCCT	CAAAAACGGG	GAATTGAGCG
801	TATCGGTTGC	GGAAAAGTAC	GAACGTTTGG	GACTGTATGC	CGATGCGGTC
851	AAATGGGTCA	AACAGCATTA	TCCGCAaAC	CGCCGCCCCG	AGCTTTTGGA
901	AGCCTTTGTC	GAAAGCGTGC	GCTTTTGGG	CGAGCGCGAA	CAGCAGAAAG
951	CCATCGATTT	TGCCGATGCT	TGGCTGAAAG	AACAGCCCGA	TAACGCGCTT
1001	CTGCTGATGT	ATCTCGGTTCG	GCTCGCCTTC	GGCCGCAAAAC	TTTGGGGCAA
1051	GGCAAAAGGC	TACCTTGAAG	CGAGCATTGC	ATTAAAGCCG	AGTATTTCCG
1101	CGCGTTTGGT	TCTAACAAAG	GTTTTCGACG	AAATCGGAGA	ACCGCAGAAG
1151	GCGGAGGCGC	AC...			

- This corresponds to the amino acid sequence <SEQ ID 750; ORF100>:

1	MKTVVWIVVL	FAAAVGLALA	SGIYTGdVYI	VLGQTMLRIN	LHAFVLGSLI
51	AVVVWYFLFK	FIIGVLNIPE	KMQRFSGSARK	GXXKXLALNK	AGLAYFEGRF
101	EKAELASRV	LVNKVGRDNR	TLALMLXAHA	AGOMENIXXR	DRYLAELAKL
151	PEKQQLSRYL	LLAESALNRR	DYEAEEANLH	AAAKMNANLT	RLVRLXIRYA
201	FDRGDALQVL	AKTEKLSKAG	ALGKSEMERY	QNWAYRRQLA	DAADAAALKT
251	CLKRIPDSLK	NGELSVSVAE	KYERLGLYAD	AVKWWKQHYP	XNRRPELLEA
301	FVESVRFLGE	REQQKAIDFA	DAWLKEQPDN	ALLLMYLGR	LAFGRKLWGKA
351	KGYLEASIAL	KPSISARLVL	TKVFEIIGEP	QKAEAH...	

Further work revealed the complete nucleotide sequence <SEQ ID 751>:

1	ATGAAAACGG	TAGTCTGGAT	TGTCGTCCTG	TTTGCCGCCG	CCGTCGGACT
51	GGCGCTGGCT	TCGGGCATTT	ACACCGGCGA	CGTGTATATC	GTA CTGGAC
101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTCGCTGATT
151	GCCGTCGTGG	TGTGGTATTT	CTTGTTTAAA	TTCATTATCG	GCGTACTCAA



-420-

201 TATCCCCGAA AAGATGCAGC GTTTCGGTTC GGCGCGTAAA GGCCGCAAGG  
 251 CCGCGCTTGC CTTGAACAAG GCGGGTTTGG CGTATTTTGA AGGGCGTTT  
 301 GAAAAGGCGG AACTAGAAGC CTCACGCGTG TTGGTCAACA AAGAGGCCGG  
 351 AGACAACCGG ACTTTGGCAT TGATGCTGGG CGCGCACGCC GCCGGACAGA  
 401 TGGAAAACAT CGAGCTGCGC GACCGTTATC TTGCGGAAAT CGCCAAACTG  
 451 CCGGAAAAAC AGCAGCTTTC CCGTTATCTT TTGTTGGCGG AATCGGCGTT  
 501 GAACCGGCGC GATTACGAAG CGGCGGAAGC CAATCTTCAT GCGGCGGCGA  
 551 AGATGAATGC CAACCTTACG CGCCTCGTGC GTCTGCAACT TCGTTACGCT  
 601 TTCGACAGGG GCGACGCGTT GCAGGTCTG GCAAAAACCG AAAAATTTTC  
 651 CAAGGCGGGC GCGTTGGGCA AATCGGAAAT GGAACGGTAT CAAAATTGGG  
 701 CATACGCGCG CCAGCTGGCG GATGCTGCCG ATGCCGCCGC TTTGAAAACC  
 751 TGCCTGAAGC GGATTCCCGA CAGCCTCAA AAGCGGGAAT TGAGCGTATC  
 801 GGTTGCGGAA AAGTACGAAC GTTTGGGACT GTATGCCGAT GCGGTCAAAT  
 851 GGGTCAAACA GCATTATCCG CACAACCGCC GCCCGAGCT TTTGGAAGCC  
 901 TTTGTCGAAA GCGTGGCTT TTTGGGCGAG CGGAACAGC AGAAAGCCAT  
 951 CGATTTTGCC GATGCTTGGC TGAAAGAACA GCCCGATAAC GCGCTTCTGC  
 1001 TGATGTATCT CGGTGCGCTC GCCTACGGCC GCAAACCTTG GGGCAAGGCA  
 1051 AAAGCTACC TTGAAGCGAG CATTGCATTA AAGCCGAGTA TTTCCGCGCG  
 1101 TTTGGTTCTA GCAAAGGTTT TCGACGAAAT CGGAGAACCG CAGAAGCGCG  
 1151 AGGCGCAGCG CAACCTGGTT TTGGAAGCCG TCTCCGATGA CGAACGTCAC  
 1201 GCAGCGTTAG AGCAGCATAG CTGA

This corresponds to the amino acid sequence <SEQ ID 752; ORF100-1>:

1 MKTVVWIVVL FAAAVGLALA SGIYTGdVYI VLgQTMLRIN LHAFVLGSLI  
 51 AVVVWYFLFK FIIGVLNIPE KMQRFGSARK GRKAALALNK AGLAYFEGRF  
 101 EKAELEASRV LVNKEAGDNR TLALMLGAHA AGQMENIELR DRYLAETIAKL  
 151 PEKQQLSRYL LLAESALNRR DYEAEEANLH AAKMNNANLT RLVRLQLRYA  
 201 FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQLA DAADAAALKT  
 251 CLKRIPDSLK NGELSVSVAE KYERLGLYAD AVKWKQHYP HNRRPELLEA  
 301 FVESVRFGE REQQAIDFA DAWLKEQPDN ALLMYLGR L AYGRKLWGKA  
 351 KGYLEASIAL KPSISARLVL AKVFDEIGEP QKAEAQRLV LEAVSDDERH  
 401 AALEQHS\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF100 shows 93.5% identity over a 386aa overlap with an ORF (ORF100a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60	
orf100.pep		MKT	VVW	IVVL	FAAAVGLALA	SGIYTGdVYI	VLgQTMLRIN	LHAFVLGSLI
orf100a		MKT	VVW	IVVL	FAAAVGLALA	SGIYTGdVYI	VLgQTMLRIN	LHAFVLGSLI
		10	20	30	40	50	60	
		70	80	90	100	110	120	
orf100.pep		FIIGVLNIPE	KMQRFGSARK	GKXKX	LLALNKAGLAY	FEGRF	EKAELEASRV	LVNKEAGDNR
orf100a		FIIGVLNXP	KMQRFGSARK	GRKAALALNK	AGLAYFEGRF	EKAELEASRV	LVNKEAGDNR	
		70	80	90	100	110	120	
		130	140	150	160	170	180	
orf100.pep		TLALMLXAHA	AGQMENIXR	DRYLAEIAKL	PEKQQLSRYL	LLAESALNRR	DYEAEEANLH	
orf100a		TLALMLGAHA	AGQMENIELR	DRYLAEIAKL	PEKQQLSRYL	LLAESALNRR	DYEAEEANLH	
		130	140	150	160	170	180	
		190	200	210	220	230	240	
orf100.pep		AAAKMNNANL	TRLVRLXIR	YAFDRGDALQ	VLAKTEKLSKAG	ALGKSEMERY	QNWAYRRQLA	
orf100a		AAAKMNNANL	TRLVRLQLR	YAFDRGDALQ	VLAKTEKLSKAG	ALGKSEMERY	QNWAYRRQLX	
		190	200	210	220	230	240	
		250	260	270	280	290	300	
orf100.pep		DAADAAALKT	CLKRIPDSL	KNGELSVSVAE	KYERLGLYAD	AVKWKQHYP	PNRRPELLEA	
orf100a		DAADAAALKT	CLKRIPDSL	KNGELSVSVAE	KYERLGLYAD	AVKWKQHYP	PNRRPELLEA	
		250	260	270	280	290	300	

-421-

		310	320	330	340	350	360
	orf100.pep	FVESVRFLGEREQQKAIDFADAWLKEQPDNALLMYLGRILAFGRKLWGKAKGYLEASIAL					
5	orf100a	:       :       :       :       :       :       :					
		310	320	330	340	350	360
	orf100.pep	KPSISARLVLTQVFDEIGEPQKAEAH	370	380			
10	orf100a	:       :       :       :       :       :					
		KPSISARLVLAQVFDETGEPEQKAEQAQRNLVLASVAEENRPSAETHX					
		370	380	390	400		

The complete length ORF100a nucleotide sequence <SEQ ID 753> is:

	1	ATGAAAACGG	TAGTCTGGAT	TGTCGTCTCTG	TTGCGCGCCG	CNNTCGGGCT
15	51	GGCATTGGCG	TCGGGCATTN	ACACCGGCGA	CGTGTATATC	GTACTCGGAC
	101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTTCGCTGATT
	151	GCCGTCGTGG	TGTGGTATTT	CCTGTTCAAA	TTCATCATCG	GCGTACTCAA
	201	TANCCCGAA	AAGATGCAGC	GTTTCGGTTC	GGCGCGTAAA	GGCCGCAAGG
	251	CCGCGCTTGC	TTGAACAAG	GCGGGTTTGG	CGTATTTTGA	AGGGCGTTTT
20	301	GAAAAGCGG	AACCTGAAGC	CTCGCGCGTA	TTGGGAAACA	AAGAGGCGGG
	351	GGATAACCGG	ACTTTGGCAT	TGATGTTGGG	CGCACATGCC	GCCGGGCAGA
	401	TGGAAAACAT	CGAGCTGCGC	GACCGTTATC	TTGCGGAAAT	CGCCAACTG
	451	CCGGAAGAGC	AGCAGCTTTC	CCGTTATCTT	TTGTTGGCGG	AATCGGCGTT
	501	GAACCGGCGC	GATTACGAAG	CGGCGGAAGC	CAATCTTCAT	GCGGCGGCGA
25	551	AGATGAATGC	CAACCTTACG	CGCCTCGTGC	GTCTGCAACT	TCGTTACGCT
	601	TTCGACAGGG	GCGACGCGTT	GCAGGTTCTG	GCAAAAACCG	AAAAANTTTC
	651	CAAGGCGGGC	GCGTNGGGCA	AATCGGAAAT	GGAACGGTAT	CAAAATTGGG
	701	CATACCGCCG	CCAGCTGNCG	GATGCTGCCG	ATGCCGCCGC	TTTGAAAACC
	751	TGCCTGAAGC	GGATTCCCAG	CAGCCTCAAA	AACGGGGAAT	TGAGCGTATC
30	801	GGTTGCGGAA	AAGTACGAAC	GTTTGGGACT	GTATGCCGAT	GCGGTCAAAT
	851	GGGTCAAACA	GCATTATCCG	CACAACCGCC	GACCCGAAT	TTTGGAAGCN
	901	TTTGTCGAAA	GCGTGCCTT	TTTGGGCGAA	CGCGATCAGC	AGAAAGCCAT
	951	CGATTTTGCC	GATGCTTGGC	TGAAAGAACA	GCCCGATAAT	GCGCTTCTGC
	1001	TGANGTATCT	CGTTCGGCTC	GCCTACGGCC	GCAAACCTTG	GGGCAAGGCA
35	1051	AAAGGCTACC	TTGAAGCGAG	CATTGCATTA	AAGCCGAGTA	TTTCCGCGCG
	1101	TTGGTTCTG	GCAAAGGTTT	TTGACGAAAC	CGGAGAACCG	CAGAAGCGCG
	1151	AGGCGCAGCG	CAACTTGSTT	TTGGCAAGCG	TTGCCGAGGA	AAACCGNCCT
	1201	TCCGCCGAAA	CCCATTGA			

This encodes a protein having amino acid sequence <SEQ ID 754>:

	1	MKTVVWIVVL	FAAAXGLALA	SGIXTGDVYI	VLGQTMLRIN	LHAFVLGSLI
40	51	AVVVWYFLFK	FIIGVLNXPE	KMQRFSGARK	GRKAALALNK	AGLAYFEGRF
	101	EKAELEASRV	LGNKEAGDNR	TLALMLGAHA	AGQMENIELR	DRYLAEIAKL
	151	PEKQQLSRYL	LLAESALNRR	DYEAEEANLH	AAAKMNANLT	RLVRLQLRYA
	201	FDRGDALQVL	AKTEKXSKAG	AXGKSEMERY	QNWAYRRQLX	DAADAAALKT
45	251	CLKRIPDSLK	NGELSVSVAE	KYERLGLYAD	AVKWKQHYP	HNRRPELLEA
	301	FVESVRFLGE	RDQQKAIDFA	DAWLKEQPDN	ALLLXYLGR	AYGRKLWGKA
	351	KGYLEASIAL	KPSISARLV	AKVFDDETGE	QKAEAQRLV	LASVAEENRP
	401	SAETH*				

ORF100a and ORF100-1 show 95.1% identity in 406 aa overlap:

		10	20	30	40	50	60
50	orf100a.pep	MKTVVWIVVLF	FAAAXGLALAS	GIXTGDVYIV	VLGQTMLRIN	LHAFVLGSLI	AVVVWYFLFK
	orf100-1	MKTVVWIVVLF	FAAAVGLALAS	GIYTGDDVYI	VLGQTMLRIN	LHAFVLGSLI	AVVVWYFLFK
		10	20	30	40	50	60
55	orf100a.pep	FIIGVLNXPE	KMQRFSGARK	GRKAALALNK	AGLAYFEGRF	EKAELEASRV	LGNKEAGDNR
	orf100-1	FIIGVLNIPE	KMQRFSGARK	GRKAALALNK	AGLAYFEGRF	EKAELEASRV	LVNKEAGDNR
		70	80	90	100	110	120
60	orf100a.pep	TLALMLGAHA	AGQMENIELR	DRYLAEIAKL	PEKQQLSRYL	LLAESALNRR	DYEAEEANLH
	orf100-1	TLALMLGAHA	AGQMENIELR	DRYLAEIAKL	PEKQQLSRYL	LLAESALNRR	DYEAEEANLH
		130	140	150	160	170	180
65	orf100a.pep	TLALMLGAHA	AGQMENIELR	DRYLAEIAKL	PEKQQLSRYL	LLAESALNRR	DYEAEEANLH
	orf100-1	TLALMLGAHA	AGQMENIELR	DRYLAEIAKL	PEKQQLSRYL	LLAESALNRR	DYEAEEANLH

-422-

		130	140	150	160	170	180
5	orf100a.pep	190	200	210	220	230	240
	orf100-1	190	200	210	220	230	240
10	orf100a.pep	250	260	270	280	290	300
	orf100-1	250	260	270	280	290	300
15	orf100a.pep	310	320	330	340	350	360
	orf100-1	310	320	330	340	350	360
20	orf100a.pep	370	380	390	400		
	orf100-1	370	380	390	400		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF100 shows 93.3% identity over a 386 aa overlap with a predicted ORF (ORF100ng) from *N.gonorrhoeae*:

30	orf100.pep	MKTVVWIVVLF	AAAVGLALASGIY	TG	DVYIVL	QTMRLINLHAF	VLGSLIAVVVWY	FLFK	60
	orf100ng	MKTVVWIVVLF	AAAVGLALASGIY	TG	DVYIVL	QTMRLINLHAF	VLGSLIAVVVWY	FLFK	60
35	orf100.pep	FIIGVLN	PIPEKMRFGS	SARKGXKXXL	ALNKA	LAYFEGR	FEKAELEAS	RVLVNKG	RDNR 120
	orf100ng	FIIGVLN	PENMRRSG	SARKGRKA	ALNKA	LAYFEGR	FEKAELEAS	RVLVNKG	RDNR 120
40	orf100.pep	TLALMLX	AHAAGOMEN	IXRDRY	LAEIAKL	PEKQQLS	RYLLLAES	ALNRRD	YEAEEANLH 180
	orf100ng	TLALMLG	AHAAGOMEN	IELRDRY	LAEIAKL	PEKQQLS	RYLLLAES	ALNRRD	YEAEEANLH 180
45	orf100.pep	AAAKMNAN	LTRLVRLX	IRYAFDR	GDALQV	LAKTEKLS	KAGALGK	SEMER	YQNWAYRRQLA 240
	orf100ng	AAAKMNAN	LTRLVRLQ	LYAFDR	GDALQV	LAKTEKLS	KAGALGK	SEMER	YQNWAYRRQMA 240
50	orf100.pep	DAADAAAL	KTCLKRIP	DSLKN	GELSVS	VAEKYER	LGLYAD	AVKWKQ	HYPHNRRPELLEA 300
	orf100ng	DAADAAAL	KTCLKRIP	DSLKN	GELSVS	VAEKYER	LGLYAD	AVKWKQ	HYPHNRRPELLEA 300
55	orf100.pep	FVESVRFL	GEREQQKA	IDFAD	AWLKEQ	PDNALL	MYLGR	LAFGR	KLWGKAKGYLEASIAL 360
	orf100ng	FVESVRFL	GEREQQKA	IDFAD	SWLKEQ	PDNALL	MYLGR	LAYGR	KLWGKAKGYLEASIAL 360
55	orf100.pep	KPSISAR	LVLTKV	FDEIGE	PQKAEAH				386
	orf100ng	KPSIPAR	LVLAKV	FDETAQ	SQKAE	AQRNLV	LASV	AGENR	PSAETR 405

The complete length ORF100ng nucleotide sequence <SEQ ID 755> is:

1	ATGAAACGG	TAGTCTGGAT	TGTTGTCCTG	TTTGCCGCCG	CCGTCGGACT
51	GGCGCTGGCT	TCGGGCATT	ACACCGGCGA	CGTGTATATC	GTACTCGGAC
101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTCGCTGATT
151	GCCGTCGTGG	TGTGGTATT	CCTGTTTAA	TTCATCATCG	GCGTACTCAA
201	TATCCCCGAA	AATATGCGGC	GTTCCGGTTC	GGCGCGGAAA	GGCCGCAAGG
251	CCGCGCTTGC	CTTGAATAAG	GCGGGTTTGG	CGTATTTCGA	AGGGCGTTT
301	GAAAGGCGG	AACTCGAAGC	CTCTCGAGTG	TTGGGCAACA	AAGAGCCCGG
351	AGACAACCGG	ACTTTGGCAT	TGATGCTGGG	CGCGCACGCG	GCAGGACAGA
401	TGGAAATAT	CGAGCTGCGC	GACCGTTATC	TTGCGGAAAT	CGCCAACTG

-423-

5  
10  
15

```

451 CCGGAAAAAC AGCAGCTTTC CCGCTATCTT CTGCTGGCGG AATCGGCGTT
501 AAACCGGCGC GATTACGAAG CGGCGGAAGC CAATCTTCAT GCGGCGGCGA
551 AGATGAATGC CAACCTTACG CGCCTCGTGC GTCTGCAACT TCGTTACGCC
601 TTCGATCGGG GCGATGCGTT GCAGGTTCTG GCAAAAaccG AAAAATTTC
651 CAAGGCGGGC GCGTTGGGCA AATCGGAAAT GGAACGGTAT CAAAATTGGG
701 CATAACGCCG CCAGATGGCG GATGCTGCCG ATGCCGCCGC TTTGAAAACC
751 TGCCTGAAGC GGATTCCCGA CAGCCTCAA AACC GGGAAT TGagcGTATC
801 GGTTCGGGAA AAGTACGAAC GTTGGGACT GTATGCCGAT GCGGTCAAAT
851 GGGTCAAAAC GCATTATCCG CACAACCGCC GCCCGAGCT TTTGGAAGCC
901 TTTGTCGAAA GCGTGCCTTT TTTGGGCGAG CGCGAACAGC AGAAAGCCAT
951 CGATTTTGCC GATTCTTGGC TGAAAGAACA GCCCGATAAC GCGCTTCTGC
1001 TGATGTATCT CGGCCGGCTC GCCTACGGCC GCAAACTTTG GGGTAAGGCA
1051 AAAGGCTACC TTGAAGCGAG TATTGCACTG AAGCCGAGTA TTCGGCGCGC
1101 TTTGGTGTG GCAAAGGTTT TTGACGAAAC CGCACAGTCG CAAAAGCCG
1151 AAGCACAGCG CAACTTGTTT TTGGCAAGCG TTGCCGGGGA AAACCGCCCT
1201 TCCGCCGAAA CCCGTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 756>:

20  
25

```

1 MKTVVWIVVL FAAAVGLALA SGIYTGdVYI VLQQTMLRIN LHAFVLGSLI
51 AVVVWYFLFK FIIGVLNIPE NMRRSGSARK GRKAALALNK AGLAYFEGRF
101 EKAELEASRV LGNKEAGDNR TLALMLGAHA AGQMENIELR DRYLAEIAKL
151 PEKQQLSRYL LLAESALNRR DYEAAEANLH AAKMNANLT RLVLRLQRYA
201 FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQMA DAADAAALKT
251 CLKRIPDSLK NGELSVSVAE KYERLGLYAD AVKWKQHYH HNRRPELLEA
301 FVESVRFLGE REQQAIDFA DSWLKEQPDN ALLMYLGR LAYGRKLWGKA
351 KGYLEASIAL KPSIPARLV AKVFDETAQS QKAEAQRLV LASVAGENRP
401 SAETR*

```

ORF100ng and ORF100-1 show 95.3% identity in 402 aa overlap:

30  
35  
40  
45  
50  
55  
60  
65

```

          10      20      30      40      50      60
orf100-1.pep MKTVVWIVVLF FAAAVGLALASGIYTGdVYI VLQQTMLRIN LHAFVLGSLI AVVVWYFLFK
          10      20      30      40      50      60
orf100ng      MKTVVWIVVLF FAAAVGLALASGIYTGdVYI VLQQTMLRIN LHAFVLGSLI AVVVWYFLFK

          70      80      90      100     110     120
orf100-1.pep FIIGVLNIPE KMQRFGSARK GRKAALALNK AGLAYFEGRF EKAELEASRV LGNKEAGDNR
          70      80      90      100     110     120
orf100ng      FIIGVLNIPENMRRSGSARK GRKAALALNK AGLAYFEGRF EKAELEASRV LGNKEAGDNR

          130     140     150     160     170     180
orf100-1.pep TLALMLGAHA AGQMENIELR DRYLAEIAKL PEKQQLSRYL LLAESALNRR DYEAAEANLH
          130     140     150     160     170     180
orf100ng      TLALMLGAHA AGQMENIELR DRYLAEIAKL PEKQQLSRYL LLAESALNRR DYEAAEANLH

          190     200     210     220     230     240
orf100-1.pep AAKMNANLTRLVRLQRYA FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQLA
          190     200     210     220     230     240
orf100ng      AAKMNANLTRLVRLQRYA FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQMA

          250     260     270     280     290     300
orf100-1.pep DAADAAALKT CLKRIPDSLK NGELSVSVAE KYERLGLYAD AVKWKQHYH PHNRRPELLEA
          250     260     270     280     290     300
orf100ng      DAADAAALKT CLKRIPDSLK NGELSVSVAE KYERLGLYAD AVKWKQHYH PHNRRPELLEA

          310     320     330     340     350     360
orf100-1.pep FVESVRFLGEREQQAIDFADAWLKEQPDN ALLMYLGR LAYGRKLWGKAKGYLEASIAL
          310     320     330     340     350     360
orf100ng      FVESVRFLGEREQQAIDFADSWLKEQPDN ALLMYLGR LAYGRKLWGKAKGYLEASIAL

          370     380     390     400
orf100-1.pep KPSISARLV LAKVFDEIGEPQKAE AQRLVLEAVSDDERHAAL EQHSX
          370     380     390     400
orf100ng      KPSIPARLV LAKVFDETAQS QKAE AQRLV LASVAGENRPSAETR*

```

370 380 390 400

Based on this analysis, including the presence of a putative leader sequence, a putative transmembrane domain, and a RGD motif, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 90

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 757>

```

10      1  ATGATGTTTT CTGGTTCAA GCTGTTTAC TTGTTTTTG TCATTTCTGTG
      51  GTTTGCAGGG CTGTTTACC TGCCGAGGAT TTTCGTCAAT ATGGCGATGA
     101  TTGATGTGCC GCGCGGCAAT CCCGAGTATG TCGCTCTGTC GGGCATGGCG
     151  GTGCGGCTGT ACCGTTTTAT GTCGCCGTTG GGCTTCGGCG CGGTCGTGTT
     201  CGGCGCGGCG ATACCGTTTG CCGCCGCTG GTGGGCGAGC GGCTGGGTAC
     251  ACGTCAAAC TGTGTTGGGC TTGATGCTCT TGGCTTACCA GTTGATTGTC
     15  301  GCGGTGCTGC TGCGCCGTT TCAGGATTAC AGCAATGCTT TTTCACACCG
     351  CTGGTACCGC GTGTTCAACG AAATCCCCGT GCTGCTGATG GTTGCCGCGC
     401  TGTATCTGGT CGTGTTCAAA CCGTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 758; ORF102>:

```

20      1  MMFSWFKLFH LFFVISWFAG LFYLPRI FVN MAMIDVPRGN PEYVRLSGMA
      51  VRLYRFMSPL GFGAVVFGAA IPFAAGWWGS GWVHVKLCLG LMLLAYQLYC
     101  GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYXVVEK PF*

```

Further work revealed the complete nucleotide sequence <SEQ ID 759>:

```

25      1  ATGATGTTTT CTGGTTCAA GCTGTTTAC TTGTTTTTG TCATTTCTGTG
      51  GTTTGCAGGG CTGTTTACC TGCCGAGGAT TTTCGTCAAT ATGGCGATGA
     101  TTGATGTGCC GCGCGGCAAT CCCGAGTATG TCGCTCTGTC GGGCATGGCG
     151  GTGCGGCTGT ACCGTTTTAT GTCGCCGTTG GGCTTCGGCG CGGTCGTGTT
     201  CGGCGCGGCG ATACCGTTTG CCGCCGCTG GTGGGCGAGC GGCTGGGTAC
     251  ACGTCAAAC TGTGTTGGGC TTGATGCTCT TGGCTTACCA GTTGATTGTC
     301  GCGGTGCTGC TGCGCCGTT TCAGGATTAC AGCAATGCTT TTTCACACCG
     30  351  CTGGTACCGC GTGTTCAACG AAATCCCCGT GCTGCTGATG GTTGCCGCGC
     401  TGTATCTGGT CGTGTTCAAA CCGTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 760; ORF102-1>:

```

35      1  MMFSWFKLFH LFFVISWFAG LFYLPRI FVN MAMIDVPRGN PEYVRLSGMA
      51  VRLYRFMSPL GFGAVVFGAA IPFAAGWWGS GWVHVKLCLG LMLLAYQLYC
     101  GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYLVVEK PF*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with HP1484 hypothetical integral membrane protein of *H. pylori* (accession number AE000647)

ORF102 and HP1484 show 33% aa identity in 143aa overlap:

```

40      orf102  3  FSWFKLFH LFFVISWFAG LFYLPRI FVN MAMIDVPRGN PEYVRLSGMA VRLYRFMSPLGF 62
      HP1484  8  FLWVKA FHVIAVISWMAALFYLPRL FVYHAENAHKKE FVG VVQIQEK--KLYSFIAS PAM 65
      orf102  63  GAVVFGAA IPFAAG---WWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDY SNAFSHRWY 119
      HP1484  66  GFTLITGILMLLIEPTLFKSGGWLHAKLALVLLLAYHFYCKKCMRELEKDPTRRNARFY 125
      orf102  120 RVFNEIPXXXXXXXXXXXXXFKPF 142
      HP1484  126 RVFNEAPTILMILIVILVVVKPF 148

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF102 shows 99.3% identity over a 142aa overlap with an ORF (ORF102a) from strain A of *N. meningitidis*:

5		10	20	30	40	50	60
	orf102.pep	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVNMAMIDVPRGNPEYVRLSGMAVRLYR	FMSPL			
	orf102a	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVNMAMIDVPRGNPEYVRLSGMAVRLYR	FMSPL			
10		10	20	30	40	50	60
	orf102.pep	GFGAVVFGAAIPFAAGWWGSGWVHV	KLCGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
	orf102a	GFGAVVFGAAIPFAAGWWGSGWVHV	KLCGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
15		70	80	90	100	110	120
	orf102.pep						
	orf102a						
20		130	140				
	orf102.pep	VFNEIPVLLMVAALYXVVEKPF	X				
	orf102a	VFNEIPVLLMVAALYLVEKPF	X				

The complete length ORF102a nucleotide sequence <SEQ ID 761> is:

	1	ATGATGTTT	CTTGGTTCAA	GCTGTTTCAC	TTGTTTTTTG	TCATTTTCGTG
25	51	GTTTGCAGGG	CTGTTTTACC	TGCCGAGGAT	TTTCGTCAAT	ATGGCGATGA
	101	TTGATGTGCC	GCGCGCAAT	CCCGAGTATG	TGCGTCTGTC	GGGCATGGCG
	151	GTGCGGCTGT	ACCGTTTTAT	GTCGCCGTG	GGCTTCGGCG	CGGTCGTGTT
	201	CGGCGCGGCG	ATACCGTTTG	CCGCCGGCTG	GTGGGCGAGC	GGCTGGGTAC
	251	ACGTCAAAC	GTGTTTGGGC	TTGATGCTCT	TGGCTTACCA	GTTGTATTGC
	301	GGCGTGCTGC	TGCGCCGTTT	TCAGGATTAC	AGCAATGCTT	TTTACACCG
30	351	CTGGTACCGC	GTGTTCAACG	AAATCCCGT	GCTGCTGATG	GTTGCCGCGC
	401	TGTATCTGGT	CGTGTTCAAA	CCGTTTTGA		

This encodes a protein having amino acid sequence <SEQ ID 762>:

	1	MMFSWFKLFH	LFFVISWFAG	LFYLPRI	FVNMAMIDVPRGN	PEYVRLSGMA
35	51	VRLYRFMSPL	GFGAVVFGAA	IPFAAGWWGS	GWVHVKLCLG	LMLLAYQLYC
	101	GVLRRFQDY	SNAFSHRWYR	VFNEIPVLLM	VAALYLVEFK	PF*

ORF102a and ORF102-1 show complete identity in 142 aa overlap:

		10	20	30	40	50	60
	orf102a.pep	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVNMAMIDVPRGNPEYVRLSGMAVRLYR	FMSPL			
40	orf102-1	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVNMAMIDVPRGNPEYVRLSGMAVRLYR	FMSPL			
		10	20	30	40	50	60
	orf102a.pep	GFGAVVFGAAIPFAAGWWGSGWVHV	KLCGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
45	orf102-1	GFGAVVFGAAIPFAAGWWGSGWVHV	KLCGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
		70	80	90	100	110	120
	orf102a.pep						
	orf102-1						
50		130	140				
	orf102a.pep	VFNEIPVLLMVAALYLVEKPF	X				
	orf102-1	VFNEIPVLLMVAALYLVEKPF	X				

55 Homology with a predicted ORF from *N.gonorrhoeae*

ORF102 shows 97.9% identity over a 142 aa overlap with a predicted ORF (ORF102ng) from *N. gonorrhoeae*:

-426-

```

orfl02.pep  MMFSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDVPRGNPEYVRLSGMAVRLYRMSPL  60
orfl02ng    MMFSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDAPRGNPEYVRLSGMAVRLYRMSPL  60

5  orfl02.pep  GFGAVVFGAAIPFAAGRWGSGWVHV KLC LGLM L LAYQLYCGVLLRRFQDYSNAF SHRWYR  120
    orfl02ng  GFGAVVFGAAIPFAAGRWGSGWVHV KLC LGLM L LAYQLYCGVLLRRFQDYSNAF SHRWYR  120

10 orfl02.pep  VFNEIPVLLMVAALYXVVF KPF  142
    orfl02ng  VFNEIPVLLMVAALYL VVF KPF  142

```

The complete length ORF102ng nucleotide sequence <SEQ ID 763> is:

```

1  ATGATGTTTT CTTGGTTCAA GCTGTTTCAC TTGTTTTTTG TCATTTCGTG
15 51  GTTTCAGGG CTGTTTTACC TGCCGAGGAT TTTCGTCAAT ATGGCGATGA
    101 TTGATGCGCC GCGCGGCAAT CCCGAGTATG TCGCCTGTC GGGGATGGCG
    151 GTGCGGTTGT ACCGTTTTAT GTCGCCTTTG GGTTCGCGC CGGTCGTGT
    201 CCGCGCGGCG ATACCGTTTG CCGCgggccc GTGGGGCagc ggctggGTTT
    251 ACGTCAAACT GTGTTTGGGC TTGATGCTCT TGGCTTATCA GTTGATTGTC
    301 GCGGTGCTGC TCGCGCTTT TCAGGATTAC AGCAATGCTT TTTACACCCG
20 351 CTGGTACCGC GTGTTCAACg aAATCCCCGT GCTGCTGATG GTTGCCGCGC
    401 TGTATCTGGT CGTGTTCAAA CCGTTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 764>:

```

1  MMFSWFKLFH LFFVISWFAG LFYLPRI FVN MAMIDAPRGN PEYVRLSGMA
25 51  VRLYRMSPL GFGAVVFGAA IPFAAGRWGS GWVHV KLC LGLM L LAYQLYC
    101 GVLRRFQDY SNAF SHRWYR VFNEIPVLLM VAALYL VVEK PF*

```

ORF102ng and ORF102-1 show 98.6% identity in 142 aa overlap:

```

10 20 30 40 50 60
orfl02-1.pep MMFSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDVPRGNPEYVRLSGMAVRLYRMSPL
30 orfl02ng    MMFSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDAPRGNPEYVRLSGMAVRLYRMSPL
    10 20 30 40 50 60

70 80 90 100 110 120
orfl02-1.pep GFGAVVFGAAIPFAAGRWGSGWVHV KLC LGLM L LAYQLYCGVLLRRFQDYSNAF SHRWYR
35 orfl02ng    GFGAVVFGAAIPFAAGRWGSGWVHV KLC LGLM L LAYQLYCGVLLRRFQDYSNAF SHRWYR
    70 80 90 100 110 120

130 140
orfl02-1.pep VFNEIPVLLMVAALYL VVF KPF
40 orfl02ng    VFNEIPVLLMVAALYL VVF KPF
    130 140

```

In addition, ORF102ng shows significant homology to a membrane protein from *H. pylori*:

```

45 gi|2314656 (AE000647) conserved hypothetical integral membrane protein
    [Helicobacter pylori] Length = 148
    Score = 79.2 bits (192), Expect = 1e-14
    Identities = 50/147 (34%), Positives = 68/147 (46%), Gaps = 13/147 (8%)

50 Query: 3  FSWFKLFHLFFVISWFAGLFYLPRI FVN MAMIDAPRGNPEYVRLSGMAVRLYRMSPLGF 62
    F W K FH+ VISW A LFYLPRI FV A + V++ +LY F++
    Sbjct: 8  FLWVKAFHVIAVISWMAALFYLPRL FVYHAENAHKKEFVG VVQIQEK--KLVSFIASPA 65

55 Query: 63  GAVVFGAAIP-----FAAGRWGSGWVHV KLC LGLM L LAYQLYCGVLLRRFQDYSNAF 115
    G + + F +G GW+H KL L ++LLAY YC +R + +
    Sbjct: 66  GFTLITGILMLLIEPTLFKSG----GWLHAKLALVLL LAYHFYCKKCMRELEKDPTRRN 121

Query: 116  HRWYRVFNEIPXXXXXXXXXXXXXFKPF 142
R+YRVFNE P KPF
60 Sbjct: 122  ARFYRVFNEAPTILMILIVLVVVKPF 148

```

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 91

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 765>:

```

5      1  ATGGCAAAAA TGATGAAATG GCGGCTGTT GCGGCGGTCG CGGCGGCAGC
      51  GGTTTGGGGC GGATGGTCTT AACTGAAGCC CGAGCCGCAC GTGCTTGATA
     101  TTACGGAAAC GGTACGGCGC GGC // .....
      //.. ATTTCGTTTA CGATTTTGTC CGAACCGGAT ACGCCGATTA AGGCGAAGCT
     101  CGACAGCGTC GACCCGGGC TGACCACGAT GTCGTCGGGC GGTTACAACA
      101  GCAGTACGGA TACGGCTTCC AATGCGGTCT ACTATTATGC CCGTTCGTTT
     151  GTGCCGAATC CGGACGGCAA ACTCGCCACG GGGATGACGA CGCAGAATAC
     201  GGTTGAAATC GACGGCGTGA AAAATGTGCT GATTATTCCG TCGCTGACCG
     251  TGAAAAATCG CGGCGGCAAG GCGTTTGTGC GCGTGTGGG TCGGACGGC
     301  AAGCGGCGCG AACCGGAAAT CCGGACCGGT ATGAGAGACA GTATGAATAC
     15  351  CGAAGTAAAA AGCGGGTTGA AAGAGGGGGA CAAAGTGGTC ATCTCCGAAA
     401  TAACCGCCGC CGAGCAACAG GAAAGCGGCG AACGCGCCCT AGGCGGCCCC
     451  CCGCGCCGAT AA
  
```

This corresponds to the amino acid sequence <SEQ ID 766; ORF85>:

```

20      1  MAKMMKWAIV AAVAAAAVWG GWS.LKPEPH VLDITETVRR G.....
      51  .....
     101  .....
     151  .....
     201  ..... I SFTILSEPDT
     251  PIKAKLDSVD PGLTTMSSGG YNSSTDASN AVYYYARSFV PNPDGKLATG
     25  301  MTTQNTVEID GVKNVLIIPS LTVKNRGGKA FVRVLGADGK AAEREIRTGM
     351  RDSMNTVEKS GLKEGDKVVI SEITAAEQQE SGERALGGPP RR*
  
```

Further work revealed the further partial nucleotide sequence <SEQ ID 767>:

```

30      1  ..GTATCGGTCG GCGCGCAGGC ATCGGGGCAG ATTAAGATAC TTTATGTCAA
      51  ACTCGGGCAA CAGGTTAAAA AGGGCGATTT GATTGCGGAA ATCAATTGCA
     101  CCTCGCAGAC CAATACGCTC AATACGGAAG AATCCAAGTT GGAAACGTAT
     151  CAGGCGAAGC TGGTGTGCGC ACAGATTGCA TTGGGCAGCG CGGAGAAGAA
     201  ATATAAGCGT CAGGCGGCGT TATGGAAGGA AAACGCGACT TCCAAGAGG
     251  ATTTGGAAAG CGCGCAGGAT GCGTTTGCCG CCGCCAAAGC CAATGTTGCC
     301  GAGCTGAAGG CTTTAATCAG ACAGAGCAAA ATTTCCATCA ATACCGCCGA
     35  351  GTCGGAATTG GGCTACACGC GCATTACCGC AACGATGGAC GGCACGGTGG
     401  TGGCGATTCT CGTGGAAGAG GGGCAGACTG TGAACGCGGC GCAGTCTACG
     451  CCGACGATTG TCCAATTGGC GAATCTGGAT ATGATGTTGA ACAAATGCA
     501  GATTGCCGAG GGCGATATTA CCAAGGTGAA GGCGGGGCAG GATATTTCGT
     551  TTACGATTTT GTCCGAACCG GATACGCCGA TTAAGGCGAA GCTCGACAGC
     40  601  GTCGACCCCG GGCTGACCAC GATGTCGTCG GCGGTTTACA ACAGCAGTAC
     651  GGATACGGCT TCCAATGCGG TCTACTATTA TGCCCGTTCG TTTGTGCCGA
     701  ATCCGGACGG CAAACTCGCC ACGGGGATGA CGACGCAGAA TACGGTTGAA
     751  ATCGACGGCG TGAAAAATGT GCTGATTATT CCGTCGCTGA CCGTGAAAAA
     801  TCGCGGCGGC AAGGCGTTG TGCGCGTGT GGGTGCAGAC GGCAAGGCGG
     45  851  CGGAACGCGA AATCCGGACC GGTATGAGAG ACAGTATGAA TACCGAAGTA
     901  AAAAGCGGGT TGAAAGAGGG GGACAAAGTG GTCATCTCCG AAATAACCGC
     951  CGCCGAGCAA CAGGAAAGCG GCGAACGCGC CCTAGGCGGC CCGCCGCGCC
    1001  GATAA
  
```

This corresponds to the amino acid sequence <SEQ ID 768; ORF85-1>:

```

50      1  ..VSVGAQASQ IKILYVKLGQ QVKKGDIAE INSTSQNTL NTEKSLEY
      51  QAKLVSAQIA LGSAEKKYKR QAALWKENAT SKEDLESAQD AFAAAKANVA
     101  ELKALIRQSK ISINTAESEL GYTRITATMD GTVVAILVEE GQTVNAAQST
     151  PTIVQLANLD MMLNKMQIAE GDITKVKAGQ DISFTILSEP DTPIKAKLDS
     201  VDPGLTTMSS GGYNSSTDTA SNAVYYARS FVPNPDGKLA TGMTTQNTVE
     55  251  IDGVKNVLII PSLTVKNRGG KAFVRVLGAD GKAAEREIRT GMRDSMNTVE
     301  KSLKEGDKV VISEITAEQ QESGERALGG PPRR*
  
```

Computer analysis of this amino acid sequence gave the following results:



Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF85 shows 87.8% identity over a 41aa overlap and 99.3% identity over a 153aa overlap with an ORF (ORF85a) from strain A of *N. meningitidis*:

```

5      orf85.pep      MAKMMKWA AVAAVAAAVWGGWS-LKPEPHVLDITETVRRG
      orf85a          MAKMMKWA AVAAVAAAVWGGWSY LKPEPQAAYITETVRRGDISRTVSATGEISPSNLVS
                        10      20      30      40      50      60
10     orf85.pep      .....ISFTILSEPDTPIKAKLDSVDPGLTTMSSG
      orf85a          TIVQLANLDMMLNKMQIAEGDITKV KAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSSG
                        210     220     230     240     250     260
15     orf85.pep      110     120     130     140     150     160
      orf85a          GYNSSTDASNAVYYYARSEFV PNPDKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGGK
20     orf85a          GYNSSTDASNAVYYYARSEFV PNPDKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGR
                        270     280     290     300     310     320
25     orf85.pep      170     180     190     200     210     220
      orf85a          AFVRVLGADGKAAEREIRTGM RDSMNTEVKSGLKEGDKV VISEITAAEQQESGERALGGP
25     orf85a          AFVRVLGADGKAAEREIRTGM RDSMNTEVKSGLKEGDKV VISEITAAEQQESGERALGGP
                        330     340     350     360     370     380
30     orf85.pep      230
      orf85a          PRRX
                        390
      orf85a          PRRX

```

The complete length ORF85a nucleotide sequence <SEQ ID 769> is:

```

35      1  ATGGCAAAA  TGATGAAATG  GCGGCTGTT  GCGGCGGTCG  CGGCGGCAGC
      51  GGT TTGGG GC  GGATGGTCTT  ATCTGAAGCC  CGAGCCGCAG  GCTGCTTATA
      101  TTACGGAAC  GGTCAGGCGC  GGCGACATCA  GCCGGACGGT  TTCTGCAACA
      151  GGGGAGATT  CGCCGTCCAA  CCTGGTATCG  GTCGGCGCGC  AGGCATCGGG
      201  GCAGATTAAG  AAAC TTATG  TCAA ACTCGG  GCAACAGGTT  AAAAAGGGCG
      251  ATTTGATTGC  GGAAATCAAT  TCGACCTCGC  AGACCAATAC  GCTCAATACG
40     301  GAAAATCCA  AATTGGAAC  GTATCAGGCG  AAGCTGGTGT  CGGCACAGAT
      351  TGCATTGGGC  AGCGCGGAGA  AGAAATATAA  GCGTCAGGCG  GCGTTGTGGA
      401  AGGATGATGC  GACC GCTAAA  GAAGATTTGG  AAAGCGCACA  GGATGCGCTT
      451  GCGCGGCCA  AAGCCAATGT  TGCCGAGCTG  AAGGCTCTAA  TCAGACAGAG
      501  CAAAATTCC  ATCAATACCG  CCGAGTCGGA  ATTGGGCTAC  ACGCGCATT
45     551  CCGCAACGAT  GGACGCGACG  GTGGTGGCGA  TTCTCGTGGA  AGAGGGGCAG
      601  ACTGTGAACG  CGGCGCAGTC  TACGCCGACG  ATTGTCCAAT  TGGCGAATCT
      651  GGATATGATG  TTGAACAAA  TGCAGATTGC  CGAGGGCGAT  ATTACCAAGG
      701  TGAAGCGGG  GCAGGATATT  TCGTTACGA  TTTTGTCCGA  ACCGGATACG
50     751  CCGATTAAG  CGAAGCTCGA  CAGCGTCGAC  CCCGGGCTGA  CCACGATGTC
      801  GTCGGGCGGC  TACAACAGCA  GTACGGATAC  GGCTTCCAAT  GCGGTCTACT
      851  ATTATGCCCG  TTCGTTTGTG  CCGAATCCGG  ACGGCAAAC  CGCCACGGGG
      901  ATGACGACGC  AGAATACGGT  TGAATCGAC  GGTGTGAAAA  ATGTGCTGAT
      951  TATTCGTCG  CTGACCGTGA  AAAATCGCG  CGGCAGGGCG  TTTGTGCGCG
55     1001  TGT TGGGTGC  AGACGGCAAG  GCGCGGAAC  GCGAAATCCG  GACCGGTATG
      1051  AGAGACAGTA  TGAATACCGA  AGTAAAAGC  GGGTTGAAAG  AGGGGGACAA
      1101  AGTGGTCATC  TCCGAATAA  CCGCGCCGA  GCAGCAGGAA  AGCGCGCAAC
      1151  GCGCCCTAGG  CGGCCGCCG  CGCGGATAA

```

This encodes a protein having amino acid sequence <SEQ ID 770>:

```

60      1  MAKMMKWA AVAAVAAVWG  GWSYLKPEPQ  AAYITETVRR  GDISRTVSAT
      51  GEISPSNLVS  VGAQASQIK  KLYVKLGQQV  KKGDLIAEIN  STSQNTLNT
      101  ESKLETYQA  KLVSAQIALG  SAEKKYKRQA  ALWKDDATAK  EDLESAQDAL
      151  AAKANVAEL  KALIRQSKIS  INTAESELGY  TRITATMDGT  VVAILVEEGQ
      201  TVNAAQSTPT  IVQLANLDM  LNMQIAEGD  ITKV KAGQDI  SFTILSEPDT
65     251  PIKAKLDSVD  PGLTTMSSGG  YNSSTDASN  AVYYYARSEF  PNPDKLATG
      301  MTTQNTVEID  GVKNVLIIPS  LTVKNRGGRA  FVRVLGADGK  AAEREIRTGM

```

351 RDSMNTTEVKS GLKEGDKVVI SEITAAEQQE SGERALGGPP RR\*

ORF85a and ORF85-1 show 98.2% identity in 334 aa overlap:

```

5      orf85a.pep  30      40      50      60      70      80
      orf85-1      VSVGAQASGQIKILYVKLGQQVKKGDLIAE
                                10      20      30

10     orf85a.pep  90      100     110     120     130     140
      orf85-1      INSTSQTNLTNTEKSKLETYQAKLVSAQIALGSAEKKYKROAALWKDDATAKEDLESAQD
                                40      50      60      70      80      90

15     orf85a.pep  150     160     170     180     190     200
      orf85-1      ALAAKANVAELKALIROQSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAAQST
                                100     110     120     130     140     150

20     orf85a.pep  210     220     230     240     250     260
      orf85-1      PTIVQLANLDMMLNKMQIAEGDITKVKGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS
                                160     170     180     190     200     210

25     orf85a.pep  270     280     290     300     310     320
      orf85-1      GGYNSSTD TASNAVYYYARSFVNPDPGKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGG
                                220     230     240     250     260     270

30     orf85a.pep  330     340     350     360     370     380
      orf85-1      RAFVRLGADGKAAEREIRTGMRDSMNTTEVKSGLKEGDKVVI SEITAAEQQESGERALGG
                                280     290     300     310     320     330

35     orf85a.pep  390
      orf85-1      PPRRX
                                ||||
                                PPRRX

```

Figure 19D shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF85a..

#### Homology with a predicted ORF from *N.gonorrhoeae*

45 ORF85 shows a high degree of identity with a predicted ORF (ORF85ng) from *N.gonorrhoeae*:

```

ORF85      1 MAKMMKWA AVAAVAAA VVGWWS.LKPEPHVLDITETVRRG..... 40
            |||||
ORF85ng     1 MAKMMKWA AVAAVAAA VVGWWSY LKPEPQAAYITEAVRRGDISRTVSAT 50

50     ORF85      .....ISFTILSEPDT 250
            |||||
ORF85ng    201 TVNAAQSTPTIVQLANLDMMLNKMQIAEGDITKVKGQDISFTILSEPDT 250

55     ORF85      251 PIKAKLDSVDPGLTTMSSGGYNSSTD TASNAVYYYARSFVNPDPGKLATG 300
            |||||
ORF85ng    251 PIKAKLDSVDPGLTTMSSGGYNSSTD TASNAVYYYARSFVNPDPGKLATG 300

60     ORF85      301 MTTQNTVEIDGVKNVLIIPSLTVKNRGGKAFVRLGADGKAAEREIRTGM 350
            |||||
ORF85ng    301 MTTQNTVEIDGVKNVLIIPSLTVKNRGGKAFVRLGADGKAVEREIRTGM 350

ORF85      152 RDSMNTTEVKSGLKEGDKVVI SEITAAEQQESGERALGGPPRR 393
            :|||
65     ORF85ng    351 KDSMNTTEVKSGLKEGDKVVI SEITAAEQQESGERALGGPPRR 393

```

The complete length ORF85ng nucleotide sequence <SEQ ID 771> is:

```

1  ATGGCAAAAA TGATGAAATG GCGGCTGTT GCGGCGGTCG CCGGCGCaac
51  GGTTTGGGGC GGATGGTCTT ATCTGAAGCC CGAACCGCAG GCTGCTTATA
101 TTACGGAaac ggTCAGGCGC GCGCATATCA GCCGGACGGT TTCCGCGACG
151 GgcgAGATTT CGCCGTCCAA CCTGGTATCG GTCGGCGCGC AGGCTTCGGG
201 GCAGATTAAA AAGCTTTATG TCAAACCTCG GCAACAGGTC AAAAAGGGCG
251 ATTTGATTGC GGAAATCAAT TCGACCACGC AGACCAACAC GATCGATATG
301 GAAAAATCCA AATTGGAaAC GTATCAGGCG AAGCTGGTGT CCGCACAGAT
351 TGCATTGGGC AGCGCGGAGA AGAAATATAA GCGTCAGGCG GCGTTGTGGA
10 401 AGGATGATGC GACCTCTAAA GAAGATTGG AAAGCGCGCA GGATGCGCTT
451 GCCCGCCCA AAGCCAATGT TGCCGAGTTG AAGGCTTTAA TCAGACAGAG
501 CAAATTTCC ATCAATACCG CCGAGTCGGA TTTGGGCTAC ACGCGCATT
551 CCGCGACGAT GGACGGCACG GTGGTGGCGA TTCCCGTGGA AGAGGGGCAG
601 ACTGTGAACG CGCGCGAGTC TACGCCGACG ATTGTCCAAT TGGCGAATCT
15 651 GGATATGATG TTGAACAAAA TGCAGATTGC CGAGGGCGAT ATTACCAAGG
701 TGAAGCGGG GCAGGATATT TCGTTTACGA TTTGTCCGA ACCGGATACG
751 CCGATTAAGG CGAAGCTCGA CAGCGTCGAC CCCGGGCTGA CCACGATGTC
801 ATCGGGCGGC TACAACAGCA GTACGGATAC GGCTTCCAAT GCGTCTATT
851 ATTATGCCCG TTCGTTTGTG CCGAATCCGG ACGGCAAACT CGCCACGGGG
20 901 ATGACGACGC AGAATACGGT TGAAATCGAC GGTGTGAAAA ATGTGTTGCT
951 TATTCCGTCG CTGACCGTGA AAAATCGCGG CGGCAAGGCG TTCGTACGCG
1001 TGTTGGGTGC GGACGGCAAG GCAGTGAAC GCGAAATCCG GACCGGTATG
1051 AAAGACAGTA TGAATACCGA AGTGAAAAGC GGGTTGAAAG AGGGGGACAA
1101 AGTGGTCATC TCCGAAATAA CCGCCGCCGA GCAGCAGGAA AGCGGCGAAC
25 1151 GCGCCCTAGG CGGCCCGCGG CGCCGATAA

```

This encodes a protein having amino acid sequence <SEQ ID 772>:

```

1  MAKMMKWA AV AAAAAVWG GWSYLKPEPQ AAYITEAVRR GDISRTVSAT
51  GEISPSNLVS VQAQASGQIK KLYVKLGQOV KKGDLIAEIN STTQNTNIDM
101 EKSKLETYQA KLVSAQIALG SAEKKYKRQA ALWKDDATSK EDLESAQDAL
30 151 AAAKANVAEL KALIRQSKIS INTAESDLGY TRITATMDGT VVAIPVEEGQ
201 TVNAAQSTFT IVQLANLDM LNKMQIAEGD ITKVKAGQDI SFTILSEPDT
251 PIKAKLDSVD PGLTTMSSGG YNSSTDASN AVYYYARFV PNPDGKLATG
301 MTTQNTVEID GVKNVLLIPS LTVKNRGGKA FVRVLGADGK AVEREIRTGM
351 KDSMNTEVKS GLKEGDKVVI SEITAEQQE SGERALGGPP RR*

```

ORF85ng and ORF85-1 show 96.1% identity in 334 aa overlap:

```

30      40      50      60      70      80
orf85ng  PQAAYITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQOVKKGDLIAE
40  orf85-1  VSVGAQASGQIKLYVKLGQOVKKGDLIAE
                        10      20      30

90      100     110     120     130     140
orf85ng  INSTTQNTNIDMEKSKLETYQAKLVSAQIALGSAEKKYKRQAALWKDDATSKEDLESAQD
45  orf85-1  INSTSQTNTLNTEKSKLETYQAKLVSAQIALGSAEKKYKRQAALWKENATSKEDLESAQD
                        40      50      60      70      80      90

150     160     170     180     190     200
orf85ng  ALAAAKANVAELKALIRQSKISINTAESDLGYTRITATMDGTVVAIPVEEGQTVNAAQST
50  orf85-1  AFAAAKANVAELKALIRQSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAAQST
                        100     110     120     130     140     150

210     220     230     240     250     260
orf85ng  PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS
55  orf85-1  PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS
                        160     170     180     190     200     210

270     280     290     300     310     320
orf85ng  GGYNSSTDASNNAVYYYARFVFPNPDGKLATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG
60  orf85-1  GGYNSSTDASNNAVYYYARFVFPNPDGKLATGMTTQNTVEIDGVKNVLIIPSLTVKNRGG
                        220     230     240     250     260     270

330     340     350     360     370     380

```

-431-

```

orf85ng      KAFVRVLGADGKAVEREIRTGMKDSMNTEVKSGLKEGDKVVIS EITAAEQQESGERALGG
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf85-1      KAFVRVLGADGKAAEREIRTGMRDSMNTEVKSGLKEGDKVVIS EITAAEQQESGERALGG
                280      290      300      310      320      330

5          390
orf85ng      PPRRX
|||||
orf85-1      PPRRX

```

10 In addition, ORF85ng shows significant homology to an *E.coli* membrane fusion protein:

gi|1787104 (AE000189) o380; 27% identical (27 gaps) to 332 residues from  
membrane fusion protein precursor, MTRC\_NEIGO SW: P43505 (412 aa) [Escherichia  
coli] Length = 380

Score = 193 bits (485), Expect = 2e-48

Identities = 120/345 (34%), Positives = 182/345 (51%), Gaps = 13/345 (3%)

```

Query: 29  PQAAITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQQVKKGDLIAE 88
P   Y T  VR GD+ ++V ATG++   V VGAQ SQ+K L V +G +VKK  L+
Sbjct: 41  PVPTYQTLIVRPGDLQQSVLATGKLDALRKVDVGAQVSGQLKTL SVAIGDKVKKDQLLG 100

```

```

Query: 89  INSTTQNTIDMEKSKLETYQAKLVSAQIALGSAEKYKRQAALWKDDATSKEXXXXXX 148
I+   N I  ++ L  +A+  A+  L  A  Y RQ  L  +  A S++
Sbjct: 101 IDFEQAENQIKEVEATLMELRAQRQQAELKLARVTSRQORLAQTKAVSQQDLDTAAT 160

```

```

Query: 149 XXXXXXXXXXXXXXXXIRQSKISINTAESDLGYTRITATMDGTVAIPVEEGQTVNAAQST 208
          I++++ S++TA+++L YTRI A M G V I  +GQTV AAQ
Sbjct: 161 EMAVKQAQIGTIDAQIKRNQASLDTAKTNLDYTRIVAPMAGEVTQITTLQGQTVIAAQA 220

```

```

Query: 209 PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPLGTTMSS 268
P I+ LA++ ML K Q++E D+  +K GQ  FT+L +P T  + ++ V P
Sbjct: 221 PNILTLADMSAMLVKAQVSEADVHLKPGQKAWFTVLGDPLTRYEGQIKDVL----- 273

```

```

Query: 269 GGYNSSTDASNAVYYYARSEFVNPNDGKLATGMTTONTVEIDGVKNVLLIPSLTVKNRGG 328
+ +  ++A++YYAR VNP+G L  MT Q  +++ VKNVL IP  + + G
Sbjct: 274 -----TPEKVNDAIFYARFEVNPNGLLRLDMTAQVHIQLTDVKNVLTIPLSALGDVP 328

```

```

Query: 329 KAFVRV-LGADGKAVEREIRTGMKDSMNTEVKSGLKEGDKVVIS E 372
+V L  +G+  ERE+  G ++  + E+  GL+ GD+VVI E
Sbjct: 329 DNRYKVKLLRNGETREREVTIGARNDDVEIVKGLEAGDEVVIGE 373

```

40 Based on this analysis, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*,  
and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF85-1 (40.4kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above.  
The products of protein expression and purification were analyzed by SDS-PAGE. Figure 19A  
shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein  
45 was used to immunise mice, whose sera were used for Western blot (Figure 19B), FACS analysis  
(Figure 19C), and ELISA (positive result). These experiments confirm that ORF85-1 is a  
surface-exposed protein, and that it is a useful immunogen.

### Example 92

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 773>:

```

50      1  ..ATTCCCGCCA CGATGACATT TGAACGCAGC GGCAATGCTT ACAAATCGT
      51  TTCGACGATT AAAGTGCCGC TATACAATAT CCGTTTCGAG TCCGGCGGTA
     101  CGGTTGTCGG CAATACCCTG CACCCTACCT ACTATAGAGA CATACGCAGG
     151  GGCAAACTGT ATGCGGAAGc CAAATTCGCC GACgGcAGCG TAACTTACGG
     201  CAAAGCGGGC GAGAGCAAAA CCGAGCAAAG CCCCAAGGCT ATGGATTGTG

```

This corresponds to the amino acid sequence <SEQ ID 774; ORF120>:

10

Further work revealed the complete nucleotide sequence <SEO ID 775>:

15

20

25

This corresponds to the amino acid sequence <SEO ID 776: ORF120-1>:

30

Computer analysis of this amino acid sequence gave the following results:

35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF120 shows 92.4% identity over a 184aa overlap with an ORF (ORF120a) from strain A of *N. meningitidis*:

40

```

                                10      20      30
orf120.pep                      IPATMTTFERSGNAYKIVSTIKVPLYNIRFE
                                |||| : || ||||| ||||| |||||
orf120a      SAAILSAAALPCAYAAGLPXSAVLHYSYSGYIPATXXXXXXXXNAXKIVSTIKVPLYNIRFE
              10      20      30      40      50      60

```

45

```

              40          50          60          70          80          90
orf120.pep  SGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAMD LFTLAWQL
              |||||
orf120a     SGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTYGKAXXXXXXQSPKAMD LFTLAWQL
              70          80          90          100         110         120

```

50

```

      100      110      120      130      140      150
orf120.pep  AANDAKLPPLGLKITNGKKLYSVGGGLNKAGTGKYSIGGVETEVEVKYRVRRGDDAVMYFFAP
            |||||
orf120a     AANDAKLPPLGLKITNGKKLYSVGGGLNKAGTGKYSIGGVETEVEVKYRVRRGDDAVMYFFAP
      130      140      150      160      170      180

```

55

-433-

```

              160      170      180
orf120.pep    SLNNIPAIQIGYTDDGKTYTLKLKSVQINGQAAKPX
              |||
orf120a       SLNNIPAIQIGYTDDGKTYTLKLKSVQINGQAAKPX
              190      200      210      220

```

The complete length ORF120a nucleotide sequence <SEQ ID 777> is:

```

1  ATGATGAAGA CTTTAAAAA TATATTTTCC GCCGCCATT TGTCCGCCG
51 CCTGCCGTC GCGTATGCGG CAGGGCTGCC CNAATCCGCC GTGCTGCACT
101 ATCCGGCAG CTACGGCATT CCCGCCACNA NNANNTNNGN ACNNGNGNC
151 AATGCTTNC AAATCGTTC GACGATTAAA GTGCCGCTAT ACAATATCCG
201 TTTCGAGTCC GCGGTACGG TTGTCGGCAA TACCCTGCAC CCTACCTACT
251 ATAGAGACAT ACGCAGGGC AAACGTATG CGGAAGCCAA ATTCGCGGAC
301 GGCAGCGTAA CCTACGGCAA AGCGNNNNN ANCNNNNNNG NGCAAAGCCC
351 CAAGGCTATG GATTGTGTC CGTTGTCNTG GCAGTTGGCG GCAAATGACG
401 CGAACTCCC CCCGGGGCTG AAAATCACCA ACGGCAAAAA ACTTTATTCC
451 GTCGGCGGTT TGAATAAGC GGTACAGGA AAATACAGCA TAGGCGCGGT
501 GGAAACCGAA GTCGTCAAAT ATCGGGTGGC GCGCGGCGAC GATGCGGTAA
551 TGTATTCTT CGCACCGTCC CTGAACAATA TTCCGGCACA AATCGGCTAT
601 ACCGACGACG GCAAAACCTA TACGCTGAAA CTCAAATCGG TGCAGATCAA
651 CGGCCAGGCA GCCAAACCGT AA

```

This encodes a protein having amino acid sequence <SEQ ID 778>:

```

1  MMKTFKNIFS AAILSAAALPC AYAAGLPXSA VLHYSGSYGI PATXXXXXXX
51 NAXKIVSTIK VPLYNIRFES GGTVVGNLH PTYYRDIRRG KLYAEAKFAD
101 GSVTYGKAXX XXXXQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS
151 VGGLNKAGTG KYSIGGVETE VVKYRVRGD DAVMYFFAPS LNNIPAIQIGY
201 TDDGKTYTLK LKSVQINGQA AKP*

```

ORF120a and ORF120-1 show 93.3% identity in 223 aa overlap:

```

              10      20      30      40      50      60
orf120a.pep    MMKTFKNIFSAAILSAALPCAYAAGLPXSAVLHYSGSYGIPATXXXXXXXXNAXKIVSTIK
              |||
orf120-1       MMKTFKNIFSAAILSAALPCAYAAGLPQSAVLHYSGSYGIPATMTFERSGNAYKIVSTIK
              10      20      30      40      50      60

              70      80      90      100     110     120
orf120a.pep    VPLYNIRFESGGTVVGNLHPTYYRDIRRGKLYAEAKFADGSVTYGKAXXXXXXQSPKAM
              |||
orf120-1       VPLYNIRFESGGTVVGNLHPTYYRDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAM
              70      80      90      100     110     120

              130     140     150     160     170     180
orf120a.pep    DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRGD
              |||
orf120-1       DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRGD
              130     140     150     160     170     180

              190     200     210     220
orf120a.pep    DAVMYFFAPSLNNIPAIQIGYTDDGKTYTLKLKSVQINGQAAKPX
              |||
orf120-1       DAVMYFFAPSLNNIPAIQIGYTDDGKTYTLKLKSVQINGQAAKPX
              190     200     210     220

```

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF120 shows 97.8% identity over 184 aa overlap with a predicted ORF (ORF120ng) from *N.gonorrhoeae*:

```

55  orf120.pep                                IPATMTFERSGNAYKIVSTIKVPLYNIRFE 30
                                         |||
orf120ng    SAAILSAALPCAYAARLPQSAVLHYSGSYGIPATMTFERSGNAYKIVSTIKVPLYNIRFE 69

orf120.pep    SGGTVVGNLHPTYYRDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAMDFTLAWQL 90
|||
orf120ng    SGGTVVGNLHPAYYKDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAMDFTLAWQL 129

```

```

orf120.pep  AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGDDAVMYFFAP 150
|||||
orf120ng    AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGDDTVTYFFAP 189
5
orf120.pep  SLNNIPAIQIGYTDDGKTYTLKLKSVQINGQAAKP 184
|||||
orf120ng    SLNNIPAIQIGYTDDGKTYTLKLKSVQINGQAAKP 223

```

The complete length ORF120ng nucleotide sequence <SEQ ID 779> is:

```

10      1  ATGATGAAGA CTTTAAAAA TATATTTTCC GCCGCCATT TGTCCGCCGC
      51  CCTGCCGTGC GCGTATGCGG CAAGGCTACC CCAATCCGCC GTGCTGCACT
101     101  ATTCGGGCAG CTACGGCATT CCCGCCACGA TGACATTGA ACGCAGCGGC
151     151  AATGCTTACA AAATCGTTTC GACGATTAAA GTGCCGCTAT ACAATATCCG
201     201  TTTTGAATCC GCGGTACGG TTGTCGGCAA TACCCTGCAC CCTGCCTACT
15      251  ATAAAGACAT ACGCAGGGGC AAAGTGTATG CGGAAGCCAA ATTCGCCGAC
301     301  GGCAGCGTAA CCTACGGCAA AGCGGGCGAG AGCAAAACCG AGCAAAAGCCC
351     351  CAAGGCTATG GATTGTGTTA CGCTTGCCTG GCAGTTGGCG GCAAATGACG
401     401  CGAAATCCC CCCGGTCTG AAAATCACCA ACGGCAAAA ACTTTATTCC
451     451  GTCGGCGGCC TGAATAAGGC GGGTACGGGA AAATACAGCA TaggCGGCGT
20      501  GGAACCGAA GTCGTCAAAT ATCGGGTGCG GCGCGGCGAC GATACGGTAA
551     551  CGTATTTCTT CGCACCGTCC CTGAACAATA TTCCGGCACA AATCGGCTAT
601     601  ACCGACGACG GCAAAACCTA TACGCTGAAG CTCAAATCGG TGCAGATCAA
651     651  CGGACAGGCC GCCAAACCGT AA

```

This encodes a protein having amino acid sequence <SEQ ID 780>:

```

25      1  MMKTFKNIFS AAILSAAALPC AYAARLPQSA VLHYSYSGYGI PATMTFERSG
      51  NAYKIVSTIK VPLYNIRFES GGTVVGNLTH PAYYKDIRRG KLYAEAKFAD
101     101  GSVTYGKAGE SKTEQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS
151     151  VVGLNKAGTG KYSIGGVETE VVKYRVRRGD DTVTYFFAPS LNNIPAIQIGY
201     201  TDDGKTYTLK LKSVQINGQA AKP*

```

30 In comparison with ORF120-1, ORF120ng shows 97.8% identity in 223 aa overlap:

```

35      10      20      30      40      50      60
orf120-1.pep MMKTFKNIFS AAILSAAALPCAYAAGLPQSAVLHYSYSGYGI PATMTFERSG NAYKIVSTIK
|||||
orf120ng     MMKTFKNIFS AAILSAAALPCAYAARLPQSAVLHYSYSGYGI PATMTFERSG NAYKIVSTIK
      10      20      30      40      50      60

      70      80      90     100     110     120
orf120-1.pep VPLYNIRFESGGT VVGNLHPTYYRDIRRGKLYAEAKFADG SVTYGKAGESKTEQSPKAM
|||||
40      70      80      90     100     110     120
orf120ng     VPLYNIRFESGGT VVGNLHPAYYKDIRRGKLYAEAKFADG SVTYGKAGESKTEQSPKAM
      70      80      90     100     110     120

      130     140     150     160     170     180
orf120-1.pep DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGD
|||||
45      130     140     150     160     170     180
orf120ng     DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGD
      130     140     150     160     170     180

      190     200     210     220
orf120-1.pep DAVMYFFAPSLNNIPAIQIGYTDDGKTYTLKLKSVQINGQAAKPX
|:|
50      190     200     210     220
orf120ng     DTVTYFFAPSLNNIPAIQIGYTDDGKTYTLKLKSVQINGQAAKPX
      190     200     210     220

```

This analysis, including the presence of a putative leader sequence in the gonococcal protein suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 93

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 781>:

-435-

1 ATGTATCGGA GGAAAGGGCG GGGCATCAAG CCGTGGATGG GTGCCGGTGC  
 51 .GCGTTTGCC GCCTTGGTCT GGCTGGTTTT CGCGCTCGGC GATACTTTGA  
 101 CTCCGTTTGC GGTTCGGCG GTGCTGGCGT ATGTATTGGA CCCTTTGGTC  
 151 GAATGGTTGC AGAAAAAGGG TTTGAACCGT GCATCCGCTT CGATGCTGT  
 5 201 GATGGTGTTT TCCTTGATTT TGTGTTGGC ATTATTGTTG ATTATCGTCC  
 251 CTATGCTGGT CGGGCAGTTC AACAATTTGG CATCGCGCCT GCCCAATTA  
 301 ATCGGTTTGA TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG  
 351 CGGATATGTG GAAATCGATC AGGCATCTAT TATTGCGTGG CTTCAGGCGC  
 401 ATACGGGAGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTTGATG  
 10 451 AGGCAGGGCG GCAATATT..

This corresponds to the amino acid sequence <SEQ ID 782; ORF121>:

1 MYRRKGRGIK PWMGAGXAFA ALVWLVFALG DTLTPFAVAA VLAYVLDPLV  
 51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF>NNLASRLPQL  
 101 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW LQAHTGELSN ALKAWFPVLM  
 15 151 RQGGNI..

Further work revealed the complete nucleotide sequence <SEQ ID 783>:

1 ATGTATCGGA GGAAAGGGCG GGGCATCAAG CCGTGGATGG GTGCCGGTGC  
 51 GGCGTTTGCC GCCTTGGTCT GGCTGGTTTT CGCGCTCGGC GATACTTTGA  
 101 CTCCGTTTGC GGTTCGGCG GTGCTGGCGT ATGTATTGGA CCCTTTGGTC  
 20 151 GAATGGTTGC AGAAAAAGGG TTTGAACCGT GCATCCGCTT CGATGCTGT  
 201 GATGGTGTTT TCCTTGATTT TGTGTTGGC ATTATTGTTG ATTATCGTCC  
 251 CTATGCTGGT CGGGCAGTTC AACAATTTGG CATCGCGCCT GCCCAATTA  
 301 ATCGGTTTGA TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG  
 351 CGGATATGTG GAAATCGATC AGGCATCTAT TATTGCGTGG CTTCAGGCGC  
 25 401 ATACGGGAGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTTGATG  
 451 AGGCAGGGCG GCAATATTGT CAGCAGTATC GGCAACCTGC TCGTGCTTCC  
 501 CTGCTGCTT TACTATTTCC TGCTGGATTG GCAGCGGTGG TCGTGGCGCA  
 551 TTGCCAAACT GGTTCGAgG CGTTTGGCCG GTGCTTATAC GCGCATTACA  
 601 GCGAATTTGA ACGAGTATT GGGCGAATTT TTGCGCGGGC AGCTTCTGGT  
 30 651 AATGCTGATT ATGGGCTTGG TTTACGGTTT GGGATTGGTG CTGGTCGGGC  
 701 TGGATTCGGG GTTTGCCATC GGTATGCTTG CCGGTATTTT GGTGTTTGTC  
 751 CCTTATCTCG GGGCGTTTAC GGGATTGCTG CTTGCCACCG TCGCCGCCTT  
 801 GCTCCAGTTC GGTTCGTGGA ACGGCATCCT ATCGGTTTGG GCGGTTTTTG  
 851 CCGTAGGACA GTTTCTCGAA AGTTTTTCA TTACGCCGAA AATCGTGGGA  
 35 901 GACCGTATCG GGCTGTCGCC GTTTTGGGTT ATCTTTTCGC TGATGGCGTT  
 951 CGGGCAGCTG ATGGGCTTTG TCGGAATGTT GCGGGGATTG CCTTTGGCCG  
 1001 CCGTAACCTT GGTCTTGCTT CCGGAGGGCG TGCAGAAATA TTTTCCCGGC  
 1051 AGTTTTTACC GGGGCAGGTA G

This corresponds to the amino acid sequence <SEQ ID 784; ORF121-1>:

40 1 MYRRKGRGIK PWMGAGAAFA ALVWLVFALG DTLTPFAVAA VLAYVLDPLV  
 51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF>NNLASRLPQL  
 101 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW LQAHTGELSN ALKAWFPVLM  
 151 RQGGNIIVSSI GNLLLLPLLL YYFLLDQWRW SCGIKLVPR RFAGAYTRIT  
 45 201 GNLNEVLGEF LRGQLVMLI MGLVYGLGLV LVGLDSGFAL GMLAGILVEF  
 251 PYLGAFTGLL LATVAALLQF GSWNGILSVW AVFAVGQFLE SFFITPKIVG  
 301 DRIGLSPFWV IFSLMAFGQL MGFVGLAGL PLAAVTLVLL REGVQKYFAG  
 351 SFYRGR\*

Computer analysis of this amino acid sequence gave the following results:

#### Homology with a predicted ORF from *N.meningitidis* (strain A)

50 ORF121 shows 98.7% identity over a 156aa overlap with an ORF (ORF121a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orf121.pep	MYRRKGRGIK	PWMGAGXAFA	ALVWLVFALG	DTLTPFAVAA	VLAYVLDPLV	EWLQKKGLNR	
55 orf121a	MYRRKGRGIK	PWMDAGAAFA	ALVWLVFALG	DTLTPFAVAA	VLAYVLDPLV	EWLQKKGLNR	
		10	20	30	40	50	60
		70	80	90	100	110	120
orf121.pep	ASASMSVMVF	SLILLALLL	IIVPMLVGQF	NNLASRLPQL	IGFMQNTLLP	PWLKNTIGGYV	



-436-

```

      |||
orfl21a  ASASMSVMVFSLLLLLALLLIIVPMLVGQFNNLASRLPOLIGFMQNTLLPWLKNTIGGYV
              70      80      90      100      110      120

5      130      140      150
orfl21.pep EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNI
      |||
orfl21a  EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW
              130      140      150      160      170      180

10     190      200      210      220      230      240
orfl21a  SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLDSGF

```

The complete length ORF121a nucleotide sequence <SEQ ID 785> is:

```

15      1  ATGTATCGGA  GGAAAGGGCG  GGGCATCAAG  CCGTGGATGG  ATGCCGGTGC
      51  GCGCTTTGCC  GCCTTGGTCT  GGCTGGTTTT  CGCGCTCGGC  GATACTTTGA
      101  CTCCTTTTGC  GGTTCGGCG  GTGCTGGCGT  ATGTATTGGA  CCCTTTGGTC
      151  GAATGTTTGC  AGAAAAAGGG  TTTGAACCGT  GCATCCGCTT  CGATGTCTGT
      201  GATGGTGTTC  TCCTTGATT  TGTGTGTGGC  ATTATTGTG  ATTATTGTCC
      251  CTATGCTGGT  CCGGCAGTTC  AACAATTGG  CATCGCGCCT  GCCCCAATTA
      301  ATCGGTTTTA  TGCAGAACAC  GCTGCTGCCG  TGGTTGAAAA  ATACAATCGG
      351  CGGATATGTG  GAAATCGATC  AGGCATCTAT  TATTGCGTGG  CTTCAGGCGC
      401  ATACGGGCGA  GTTGAGCAAC  GCGCTTAAGG  CGTGGTTTCC  CGTTTTGATG
      451  AGGCAGGGCG  GCAATATTGT  CAGCAGTATC  GGCAACCTGC  TGCTGCTTCC
      501  CTTGCTGCTT  TACTATTTCC  TGCTGGATTG  GCAGCGGTGG  TCGTGCGGCA
      551  TTGCCAAACT  GGTTCGAGG  CGTTTTGCCG  GTGCTTATAC  GCGCATTACA
      601  GGCAATTTGA  ACGAGGTATT  GGGCGAATTT  TTGCGCGGGC  AGCTTCTGGT
      651  GATGCTGATT  ATGGGTTTGG  TTTACGGCTT  GGGGTTGGTG  CTGGTCGGGC
      701  TGGATTCGGG  GTTGCAATC  GGTATGGTTG  CCGGTATTTT  GGTTTTTGTT
      751  CCCTATTTGG  GCGCGTTTAC  AGGACTGCTG  CTGGCAACCG  TCGCCGCCTT
      801  GCTCCAGTTC  GGTTCGTGGA  ACGGCATCTT  GGCTGTTTGG  GCGGTTTTTG
      851  CCGTAGGACA  GTTCTCGAA  AGTTTTTCA  TTACGCCGAA  AATCGTGGGA
      901  GACCGTATCG  GCCTGTCGCC  GTTTTGGGTT  ATCTTTTCGC  TGATGGCGTT
      951  CGGGCAGCTG  ATGGGCTTTG  TCGGAATGTT  GGCCGGATTG  CCTTTGGCCG
      1001  CCGTAACCTT  GGTCTTGCTT  CGCGAGGGCG  TGCAGAAATA  TTTTGGCCGC
      1051  AGTTTTTACC  GGGGCAGGTA  G

```

This encodes a protein having amino acid sequence <SEQ ID 786>:

```

40      1  MYRRKGRGIK  PWMDAGAAFA  ALVWLVFALG  DTLTPFAVA  VLAYVLDPLV
      51  EWLQKKGLNR  ASASMSVMVF  SLILLALLL  IIVPMLVGQF  NNLASRLPOL
      101  IGFMQNTLLP  WLKNTIGGYV  EIDQASIIAW  LQAHTGELSN  ALKAWFPVLM
      151  RQGGNIVSSI  GNLLLLPLLL  YYFLLDWQRW  SCGIAKLVPR  RFAGAYTRIT
      201  GNLNEVLGEF  LRGQLLVMLI  MGLVYGLGLV  LVGLDSGF  GMVAGILVEF
      251  PYLGAFTGLL  LATVAALLQF  GSWNGILAVW  AVFAVGQFLE  SFFITPKIVG
      301  DRIGLSPFWV  IFSLMAFGQL  MGFVGMLAGL  PLAAVTLVLL  REGVQKYFAG
      351  SFYRGR*

```

45 ORF121a and ORF121-1 show 99.2% identity in 356 aa overlap:

```

      10      20      30      40      50      60
orfl21a.pep MYRRKGRGIKPWMDAGAAFAALVWLVFALGDTLTPFAVAAVLAYVLDPLVEWLQKKGLNR
      |||
orfl21-1  MYRRKGRGIKPWMDAGAAFAALVWLVFALGDTLTPFAVAAVLAYVLDPLVEWLQKKGLNR
      10      20      30      40      50      60

      70      80      90      100      110      120
orfl21a.pep ASASMSVMVFSLLLLLALLLIIVPMLVGQFNNLASRLPOLIGFMQNTLLPWLKNTIGGYV
      |||
orfl21-1  ASASMSVMVFSLLLLLALLLIIVPMLVGQFNNLASRLPOLIGFMQNTLLPWLKNTIGGYV
      70      80      90      100      110      120

      130      140      150      160      170      180
orfl21a.pep EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW
      |||
orfl21-1  EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW
      130      140      150      160      170      180

      190      200      210      220      230      240
orfl21a.pep SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLDSGF

```

[illegible]

ORF121 shows 97.4% identity over a 156 aa overlap with a predicted ORF (ORF121ng) from *N.gonorrhoeae*:

20	orf121.pep	MYRRKGRGIKPWMGAGXAFAALVWLVFALGDTLTFFAVAANLAYVLDPLVEVLQKKGLNR      :	60
	orf121.ng	MYRRKGRGIKPWMGAGAAFAALVWLVYALGDTLTFFAVAANLAYVLDPLVEVLQKKGLNR	60
25	orf121.pep	ASASMSVMVFSLILLALLLIIVPLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV 	120
	orf121.ng	ASASMSVMVFSLILLALLLIIVPLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV	120
30	orf121.pep	EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNI 	156
	orf121.ng	EIDQASIIAWFQAHTGELSNALKAWFPVLMKQGGNIVSTIGNLLLPPLLLYYFLDWHRW	180

35

1	<u>MYRRKGRGIK</u>	<u>PWMGAGAAFA</u>	ALVVLVYALG	DTLTFFAVAA	VLAYVLDPLV
51	EWLQKKGLNR	ASASMSVMVF	SLILLALLL	IIIVPLVQGQF	NNLASRLPQL
101	IGFMQNTLLP	WLKNTIGGYV	EIDQASITAW	FQAHGTGELS	ALKNAFPVLM
151	KQGGNIVSTI	GNLLLPFLL	YFLLDWHRW	SCGIPKLVPR	RFAGAYTRIT
201	GNLNKVVWKF	LRGQLLGETE	RGAVVCRVGR	ECWEGGGARS	RPSDDGWPRW
251	GGG*				

40	1	ATGTATCGGA	GAAAAGGACG	GGGCATCAAG	CCGTGGATGG	GTGCCGGCGC
	51	GGCGTTTGCC	GCCTTGGTCT	GGCTGGTTTA	CGCGCTCGGC	GATACTTTGA
	101	CTCGTTTGCG	GGTTGCGGCG	GTGCTGGCGT	ATGTGTTGGA	CCCTTTGGTG
	151	GAATGGTTGC	AGAAAAAGGG	TTTGAACCGT	GCATCCGCTT	CGATGTCTGT
	201	GATGGTGTTT	TCCTTGATTT	TGTTGTGTGC	CATATTGTTG	ATTATTGTCC
45	251	CTATGCTGGT	CGGGCAGTTC	AATAAATTGG	ATTCTCGCCT	GCCCCAATTA
	301	ATCGGTTTTA	TGCAGAACAC	GCTGCTGCCG	TGGTTGAAAA	ATACAATCGG
	351	CGGATATGTG	GAAATCGATC	AGGCATCTAT	TATTGCGTGG	TTTCAGGCGC
	401	ATACGGGCGA	GTTGAGCAAC	GCGCTTAAGG	CGTGGTTTCC	CGTTTTGATG
	451	AAACAGGGCG	GCAATATTGT	CAGCAGTATC	GGCAACCTGC	TGCTGCCGCC
50	501	CTTGCTGCTT	TACTATTTC	TGCTGGATTG	GCACGGTGG	TCGTGCCGCA
	551	TCGCCAAACT	GGTTCCGAGG	CGTTTTGCCG	GTGCTTATAC	GCGCATTACG
	601	GGTAATTTGA	ACGAGGTATT	GGCGGAATTT	TTGCGCGGTC	AGCTTCTGGT
	651	GATGCTGATT	ATGGGCTTGG	TTTACGGTTT	GGGATTGATT	CTAGTCGGAC
	701	TGGATTCCGG	ATTTGCCATC	GGTATGGTTG	CCGGTATTGT	GGTGTTTGTC
55	751	CCCTATTTGG	GTGCGTTTAC	GGGATTGCTG	CTTGCCACTG	TTGCGACCTT
	801	GCTCCAGTTC	GTTTCGTGGA	ACGGAATCTT	GGCTGTTTGG	GCGGTTTTTG
	851	CCGTCGGTCA	GTTTCTCGAA	AGTTTTTTCA	TTACGCCGAA	AATTGTAGGA
	901	GACCGTATCG	GCCTGTGCGC	GTTTTGGGTT	ATCTTTTCGC	TGATGGCGTT
	951	CGGAGAGCTG	ATGGGCTTTG	TCGGAATGTT	GGCCGGATTG	CCTTTGGCCG
60	1001	CCGTAAACCT	GGTCTTGCTT	CGCGAGGGCG	CGCAGAAATA	TTTTGCCGGC
	1051	AGTTTTTACC	GGGGCAGGTA	G		

This corresponds to the amino acid sequence <SEQ ID 790; ORF121ng-1>:

	1	MYRRKGRGIK	PWMGAGAAFA	ALVWL	VYALG	DTLTPFAVAA	VLAYVLDPLV	
	51	EWLQKKGLNR	ASASMSVMVF	SLILLALLL	IIIVPMLVGQF	NNLASRLPQL		
5	101	IGFMQNTLLP	WLKNTIGGYV	EIDQASIIAW	FQAHTGELSN	ALKAWFPVLM		
	151	KQGGNIVSSI	GNLLPPLLL	YYFLLDWQRW	SCGIAKLVPR	RFAGAYTRIT		
	201	GNLNEVLGEF	LRGQLLVMLI	MGLVYGLGLM	LVGLDSGFAI	GMVAGILVEFV		
	251	PYLGAFTGLL	LATVAALLQF	GSWNGILAVW	AVFAVGQFLE	SFFITPKIVG		
	301	DRIGLSPFWV	IFSLMAFGEL	MGFVGMLAGL	PLAAVTLVLL	REGAQKYFAG		
	351	SFYRGR*						
10	ORF121ng-1 and ORF121-1 show 97.5% identity in 356 aa overlap:							
		10	20	30	40	50	60	
	orf121-1.pep	MYRRKGRGIK	PWMGAGAAFA	ALVWL	VYALG	DTLTPFAVAA	VLAYVLDPLV	EWLQKKGLNR
	orf121ng-1	MYRRKGRGIK	PWMGAGAAFA	ALVWL	VYALG	DTLTPFAVAA	VLAYVLDPLV	EWLQKKGLNR
15		10	20	30	40	50	60	
		70	80	90	100	110	120	
	orf121-1.pep	ASASMSVMVF	SLILLALLL	IIIVPMLVGQF	NNLASRLPQL	IGFMQNTLLP	WLKNTIGGYV	
20	orf121ng-1	ASASMSVMVF	SLILLALLL	IIIVPMLVGQF	NNLASRLPQL	IGFMQNTLLP	WLKNTIGGYV	
		70	80	90	100	110	120	
		130	140	150	160	170	180	
25	orf121-1.pep	EIDQASIIAW	LQAHTGELSN	ALKAWFPVLM	RQGGNIVSS	IGNLLLP	PLLLYYFLLDWQRW	
	orf121ng-1	EIDQASIIAW	LQAHTGELSN	ALKAWFPVLM	RQGGNIVSS	IGNLLLP	PLLLYYFLLDWQRW	
		130	140	150	160	170	180	
30		190	200	210	220	230	240	
	orf121-1.pep	SCGIAKLVPR	RFAGAYTRIT	GNLNEVLGEF	LRGQLLVML	IMGLVYGLGL	MLVGLDSGFAI	
	orf121ng-1	SCGIAKLVPR	RFAGAYTRIT	GNLNEVLGEF	LRGQLLVML	IMGLVYGLGL	MLVGLDSGFAI	
		190	200	210	220	230	240	
35		250	260	270	280	290	300	
	orf121-1.pep	GMLAGILVF	PYPYLGAFT	GLLLATVAALL	QFGSWNGIL	SVWAVFAVG	QFLESFFITPKIVG	
	orf121ng-1	GMVAGILVF	PYPYLGAFT	GLLLATVAALL	QFGSWNGIL	AVWAVFAVG	QFLESFFITPKIVG	
40		250	260	270	280	290	300	
		310	320	330	340	350		
	orf121-1.pep	DRIGLSPFWV	IFSLMAFGEL	MGFVGMLAGL	PLAAVTLVLL	REGVQKYFAGS	SFYRGRX	
45	orf121ng-1	DRIGLSPFWV	IFSLMAFGEL	MGFVGMLAGL	PLAAVTLVLL	REGAQKYFAGS	SFYRGRX	
		310	320	330	340	350		

In addition, ORF121ng-1 shows homology to a permease from *H. influenzae*:

	sp P43969 PERM_HAEIN	PUTATIVE PERMEASE	PERM	HOMOLOG	Length = 349
	Score = 69.9 bits (168),	Expect = 2e-11			
	Identities = 67/317 (21%),	Positives = 120/317 (37%),	Gaps = 7/317 (2%)		
50	Query: 26	VYALGDTLTPFAVAAVLAYVLDPLVEWL-QKKGLNRASASMSVMVF	XXXXXXXXXXXXXVP	84	
		+Y GD + P +A VL+Y+L+ + +L Q R A++ +	VP		
	Sbjct: 32	IYFFGDLIAPLLIALVLSYLLIPIFNFLNQYLKCPRLATILIFGSFIGLA	AVFVLVLP	91	
55	Query: 85	MLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYVE-IDQASIIAWFQAHTGELSNALK	143		
		ML Q +L S LP + N WL N Y E ID + + + F + ++ +			
	Sbjct: 92	MLWNQTISLLSDLPAMF----NKSNEWLLNLPKNYPELIDYSMVDSIFNSVREKILGFE	147		
	Query: 144	AWFPVLMKQGGNIVSSIGNXXXXXXXXXXXXXDWQRWSCGI	AKLVPRRFAGAYTRITGNL	203	
60		+ + + N+VS D G+++ +P+ A+ R +			
	Sbjct: 148	SAVKLSLASIMNLVSLGIYAFLVPLMMFFMLKDKSELLQGVSRFLPKNRNLA	FXRWK-EM	206	
	Query: 204	NEVLGEFLRGQXXXXXXXXXXXXXXXXXDSGFAIGMVAGILVFV	PYXXXXXXXXXXXXX	263	
		+ + ++ G+ + + G+ V VPY			
65	Sbjct: 207	QQQISNYIHGKLEILIVTLITYIIFLIFGLNYP	LLLAFAVGLSVLPYIGAVIVTIPVA	266	

Query: 264 XXXXXQFGSWNGILAVWAVFAVGQFLESFFITPKIVGDRIGLSPFWVIFSLMAFGELMGF 323  
 QFG + FAV Q L+ + P + + + L P +I S++ FG L GF  
 Sbjct: 267 LVALFQFGISPTFWYIIIAFAVSQLLDGNLLVPYLFSEAVNLHPLIIISVLIFGGLWGF 326

5 Query: 324 VGMLAGLEPLAAVTLVLL 340  
 G+ +PLA + ++  
 Sbjct: 327 WGVFFAIPLATLVKAVI 343

Based on this analysis, including the presence of a putative leader sequence and transmembrane domains in the two proteins, it is predicted that the proteins from *N.meningitidis* and  
 10 *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 94

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 791>:

15 1 . .ACTGCTTTT CGGCGGCGCT GCGCTTGAGT CCATCATGAC TCGTCATATT  
 51 TTTGTCCTTT GGGAAACCGT ATCAACAAAC AGCGCCCATC TTAACATTTT  
 101 TTTGCACGTC CTGCCC GCCG CGTTCAAATG CGTACCAGCA ATACCGCCGC  
 151 CTGCGCCTCT ATGCCTTCCA TCCGCCGAG ATAGCCGAGT TTTTCGTTGG  
 201 TTTGCTCTT GATGTTGACG CACGAAATGT CTATGCCCAA ATCGCGCGCG  
 251 ATGTTGGCAC GCATTTGCGG AATGTGCGGC GCGAGTGTGG GTTCTGTGC  
 20 AATCACGGTC GTATCGACAT TGACCGCTG CCAACCCTGC GCCTGAACGC  
 351 TTTGATACGC CGCACGCAA AGGACGCGGC TGCCGCATC TTTGAACTCT  
 401 GCGGCGGTGT CCGGGAAATG GCTGCCGATA TCGCCCAAAC CTGCCGCACC  
 451 GAGCAGCGCG TCGGTAACGG CGTGACGAG CGCATCGGCA TCGGAGTGTC  
 501 CGAGCAGCCC TTTTTCAAAT GGGATTTCAG CTCCGCCAAG TATCAG..

25 This corresponds to the amino acid sequence <SEQ ID 792; ORF122>:

1 . .TAFSAALRLS PSXLVIFLSF GKPYYQTAAI LTFECTSCPP RSNAYQQYRR  
 51 LRLYAFHPPE IAEFFVGFAF DVDARNVYAO IGGDVGTHLR NVRREGFLC  
 101 NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGGVGEM AADIAQTCRT  
 151 EQRVGNVQO RIGIGVSEQP FFKWDFNSAK YQ..

30 Further work revealed the complete nucleotide sequence <SEQ ID 793>:

1 ATATCGTACT GGGCAAGCAG TTCGCCGAT TTTTGGGAAG TAGATACCGC  
 51 GCCTTTGATT TTTTGGCGC TCTTACCAA GGCTTCGATG AAAAAGTTGA  
 101 TGGTCGAGCC GGTACCGATG CCGATATATT CATTTTCGGG TACGAATTCC  
 151 ACTGCTTTT CGCGGCCGAT GCGCTTGAGT TCGTCTTGTG TCGTCATATT  
 35 TTTGCTCTT GGGAAACCGT ATCAACAAAC AGCGCCCATC TTAACATTTT  
 251 TTTGCACGTC CTGCCC GCCG CGTTCAAATG CGTACCAGCA ATACCGCCGC  
 301 CTGCGCCTCT ATGCCTTCCA TCCGCCGAG ATAGCCGAGT TTTTCGTTGG  
 351 TTTTGCCTTT GATGTTGACG CACGAAATGT CTATGCCCAA ATCGCGCGCG  
 401 ATGTTGGCAC GCATTTGCGG AATGTGCGGC GCGAGTTTGG GTTCTGTGC  
 40 AATCACGGTC GTATCGACAT TGACCGCTG CCAACCCTGC GCCTGAACGC  
 501 TTTGATACGC CGCACGCAA AGGACGCGGC TGCCGCATC TTTGAACTCT  
 551 GCGGCGGTGT CCGGGAAATG GCTGCCGATA TCGCCCAAAC CTGCCGCACC  
 601 GAGCAGCGCG TCGGTAACGG CGTGACGAG CGCATCGGCA TCGGAGTGTC  
 651 CGAGCAGCCC TTTTTCAAAT GGGATTTCAG CTCCGCCAAG TATCAGCTTT  
 45 701 CTGCTTTCG TCACTTGGTG GACATCGTAG CCTGTCCGA TACGGATGT  
 751 CGTCATCGTT TGTGTTCTCTG A

This corresponds to the amino acid sequence <SEQ ID 794; ORF122-1>:

50 1 ISYWASSSPD FLEVDTAPLI FLPLLPKASM KKLMEVPVPM PIYSFSGTNS  
 51 TAFSAAMRLS SSCVVFISF GKPYYQTAAI LTFECTSCPP RSNAYQQYRR  
 101 LRLYAFHPPE IAEFFVGFAF DVDARNVYAO IGGDVGTHLR NVRREGFLC  
 151 NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGGVGEM AADIAQTCRT  
 201 EQRVGNVQO RIGIGVSEQP FFKWDFNSAK YQLSAFGQLV DIVALSPTDV  
 251 RHRLCS\*

Computer analysis of this amino acid sequence gave the following results:

```

    5      orf122.pep                                     TAFSAALRLSPSXLVIFLSFGKPYQQTAAI  

                                                    |||||:||| |:|||||||  

    orf122a     FLPLLPKASMKKLMEVPVMPPMYSFSGTNSTAFSAAMRLSSSCVVIFLSFGKPYQQTAAI  

                    30          40          50          60          70          80  

10         40          50          60          70          80          90  

    orf122.pep     LTFFCTSCPPRSNAQQYRRLRLRYAFHPPEIAEFFVGFAFDVDARNVYAQIGGDVGTHLR  

                ||||| :||| |::|:  

    orf122a       LTFFTSCPPRSNPYQQYRRLRLRYAFHAPETIEFFVGFAFXVDARNVYAQIGGDVGTHLR  

                   90        100        110        120        130        140  

15               100        110        120        130        140        150  

    orf122.pep     NVRRECGFLCNHGRIDIDLRLPTLRLNALIRRTQKDAAVRI FELCGGVGEMAADIAQCRT  

                 |::| :|||||  

    orf122a       NMRRFEGFLCNHGRIDIDLRLPTLRLNALIRRTQKDAAVRI FELCGGVGEMAADIAQCRT  

                  150        160        170        180        190        200  

                           160        170        180  

    orf122.pep     EQRVGNVGQQRIGIGVSEQPFCKWDFNSAKYQ  

                ||||| :||| |  

25   orf122a       EQRVGNVGQQRIGIGVSEQPFCKWDFNSAKYQLSAFGQLVDIVALSDTDVHRHRLCSX  

                       210        220        230        240        250
```

	1	ATATCATATT	GGGCAAGCAG	TTCAC TGGAT	TTTTTGAAG	TAGATACCGC
30	51	GCCTTTGATT	TTTTTGCCGC	TCTTACCCAA	GGCTTCGATG	AAAAAGTTGA
	101	TGGTCGAACC	GGTACCGATG	CCGATGTATT	CGTTTTCCGG	TACGAATTCG
	151	ACTGCNTTTT	CGGCGGCGAT	GCCTTGAGT	TCGTCTTG	TCGTATATT
	201	TTTGTCCTTT	GGGAAACCGT	ATCAACAAAC	AGCCGCGATC	TTAACATTTT
35	251	TTNNNACGTC	CTGCCCGCCG	CGTTCAAATC	CTTACCAGCA	ATACCGCCGC
	301	CTGCGACTCT	ATGCCTTTCCA	TGCGCCCGAG	ATAACCGAGT	TTTTCGTTGG
	351	TTTTGCCTTT	GANGTTTGACG	CACGAAATGT	CTATGCCCAA	ATCGCGCGGC
	401	ATGTTGGCAC	GCATTTGCGG	AATATGCGGC	GCGAGTTTGG	GTTCCTGTGC
40	451	AATCAGCGTG	GTATCGACAT	TGACCGCCTG	CCAACCCCTG	GCCTGAACGC
	501	TTTGATACGC	CGCACGCAAA	AGGACGCGGC	TGTCCGCAATC	TTTGAAC TCT
	551	GCGCGCGTGT	CGGGGAAATG	GCTGCCGATA	TGCGCCAAAC	CTCGCGCACC
	601	GAGCAGCGCG	TCGGTAACGG	CGTGCAGCAG	CGCATCGGCA	TCGGAGTGTC
	651	CGAGCAGCCC	TTTTTCAAAT	GGGATTTCAA	CTCCGCCAAG	TATCAGCTTT
	701	CTGCCTTCGG	TCAGTTGGTG	GACATCGTAG	CCCTGTCCGA	TACGGATGTT
	751	CGTCATCGTT	TGTGTTCTCG	A		

45	1	ISYWASSSLD	FLEVDTAPLI	FLPLLPKASM	KKLMVEPVPM	PMYSFSGTNS
	51	TAFSAAMRLS	SSCVVIFLSF	GKPYQQTAAI	LTFFXTSCPP	RSNPYQQYRR
	101	LRLYAFHAPE	ITEFFVGFAF	XVDARNVYAQ	IGGDVGTHLR	NMRREFGFLC
	151	NHGRIDIDRL	PTLRINALIR	RTQKDAAVRI	FELCGVGEM	AADIAQTCRT
	201	EQRVGNGVQQ	RIGIGVSEQP	FFKWFDFNSAK	YQLSAFGQLV	DIVALSDDTV
50	251	RHRLCS*				

```

55
    10      20      30      40      50      60
orfl22a.pep ISYWASSSLDFLEVD TAPLIFLPLLPKASMKKLMVEPVMPMPYFSFGTNSTAFSAAMRLS
            |||||
orfl22-1    ISYWASSSPDFLEVD TAPLIFLPLLPKASMKKLMVEPVMPPIYFSFGTNSTAFSAAMRLS
            |||||

60
    70      80      90      100     110     120
orfl22a.pep SSCVVIFLSFGKPYQQTAAILTFEFTSCPPRSNPYQQYRRLRLYAFHAPETEFVGFAG
            |||||
orfl22-1    SSCVVIFLSFGKPYQQTAAILTFECTSCPPRSNAYQQYRRLRLYAFHPPEIAEFVGFAG
            |||||

```

-441-

		70	80	90	100	110	120
5	orf122a.pep	130	140	150	160	170	180
		XVDARNVYAQIGGDVGTHLRNMRREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRI					
	orf122-1	DVDARNVYAQIGGDVGTHLRNVRREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRI					
		130	140	150	160	170	180
10	orf122a.pep	190	200	210	220	230	240
		FELCGGVGEMAADIAQTCRTEQRVGNVQQRIGIGVSEQPFFKWFDFNSAKYQLSAFGQLV					
	orf122-1	FELCGGVGEMAADIAQTCRTEQRVGNVQQRIGIGVSEQPFFKWFDFNSAKYQLSAFGQLV					
		190	200	210	220	230	240
15	orf122a.pep	250					
		DIVALSDTDVRHRLCSX					
	orf122-1	DIVALSDTDVRHRLCSX					
		250					

Homology with a predicted ORF from *N.gonorrhoeae*

ORF122 shows 89.6% identity over a 182 aa overlap with a predicted ORF (ORF122ng) from *N.gonorrhoeae*:

25	orf122.pep	TAFSAALRLSPSXLVIFLSFGKPYQQTAAI	30
	orf122ng	FLPLLKASMKKLMVEPVPMPMYSFSGTNSTAFSAAMRLSSSCVVIFLSFGKPYQQTAAI	80
	orf122.pep	LTFFCTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAFDVDARNVYAQIGGDVGTHLR	90
30	orf122ng	LTFFCTSWPPRSNPYQQYRRLRLYAFHPPEIAEFFVGFAFDIDARNIDTQIGGDVGTHLR	140
	orf122.pep	NVRRECGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRIFELCGGVGEMAADIAQTCRT	150
	orf122ng	NVRCEFGFLCNHGRIDIDHPLTLRLNALIRRTQKDAAVRIFELCGGVGKMAADVAQTCRT	200
35	orf122.pep	EQRVGNVQQRIGIGVSEQPFFKWFDFNSAKYQ	182
	orf122ng	EQRVGNVQQRVGIRMPEQPFFKWFDFNSAKYQLSAFGQLVDIVALSDTDIRHRLCS	256

The complete length ORF122ng nucleotide sequence <SEQ ID 797> is:

40	1	ATGTCGTACC	GGGCAAGCAG	TTCGCCGAT	TTTTTGGAGG	TGAAACCGC
	51	GCCTTTGATT	TTTTTACCGC	TTTGGCCAA	GGCTTCGATG	AAGAAATTGa
	101	tgGTCGAACC	GgtaCCGATG	CCGATGTATT	CGTTTTCGGG	TACGAATTCG
	151	ACTGCTTTT	CGGCGCGAT	GCGCttgAgt	TCgtcttgcg	TcgTCATATT
	201	TTTAtccttt	gGGAaaccct	atcaAcaAAc	agccgccatC	TAAACATTTT
45	251	TTGTCACGtc	ctggccgcgcg	cgttcaAAc	cgtaccaGca	ataccgcccgc
	301	ctgcgcctCT	AtgcCTTCCA	TCCGCCCGAG	ATAGCCGAGT	TTTCGTTGG
	351	TTTTGCCTTT	GATatTGACG	CACGAAATAT	CGatacCCa	atcggcgGCG
	401	ATGTTGGCAC	GCATTGCGG	AATGTGCGGT	GCGAGTTTGG	GTTTCTGTGC
	451	AATCACGGTC	GTATCGACAT	TGACCACCTG	CCAACCCTGC	GCCTGAACGC
50	501	TTTGATACGC	CGCACGCAAA	AGGACGCGGC	TGTCCGCATC	TTTGAACCTC
	551	GCGGCGGTGT	CGGGAATAATG	GCTGCCGATG	TCGCCCAAAC	CTGCCGCACC
	601	GAGCAGCgcg	tcggtaaCGG	CGTGCAGCAG	cgcgTcgGCA	TCCGAATGCC
	651	CGAGCAGCCC	TTTTTCAAAT	GGGATTTCAA	CTCCGCCAAG	TATCAGCTTT
	701	CTGCCTTCGG	TCAATTGGTG	GACATCGTAG	CCCTGTCCGA	TACGGATATT
55	751	CGTCATCGTT	TGTGTTCTCTG	A		

This encodes a protein having amino acid sequence <SEQ ID 798>:

	1	MSYRASSSPD	FLEVETAPLI	FLPLLKASM	KKLMVEPVPM	PMYSFSGTNS
	51	TAFSAAMRLS	SSCVVIFLSF	GKPYQQTAAI	LTFFCTSWPP	RSNPYQQYRR
60	101	LRLYAFHPPE	IAEFFVGFAF	DIDARNIDTQ	IGGDVGTHLR	NVRCEFGFLC
	151	NHGRIDIDHL	PTLRLNALIR	RTQKDAAVRI	FELCGGVGKM	AADVAQTCRT
	201	EQRVGNVQQR	RVGIRMPEQP	FFKWFDFNSAK	YQLSAFGQLV	DIVALSDTD
	251	RHRLCS*				

ORF122ng and ORF122-1 show 92.6% identity in 256 aa overlap:

		10	20	30	40	50	60
	orf122-1.pep	ISYWASSSPDFLEVDTAPLIFLPLLPKASMKKLMVEPVPMPIYSFSGTNSTAFSAAMRLS					
5	orf122ng	MSYRASSSPDFLEVETAPLIFLPLLPKASMKKLMVEPVMPMYSFSGTNSTAFSAAMRLS					
		10	20	30	40	50	60
		70	80	90	100	110	120
10	orf122-1.pep	SSCVVIFLSFGKPYQQTAAILTFCTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAF					
	orf122ng	SSCVVIFLSFGKPYQQTAAILTFCTSWPPRSNPYQQYRRLRLYAFHPPEIAEFFVGFAF					
		70	80	90	100	110	120
		130	140	150	160	170	180
15	orf122-1.pep	DVDARNVYAQIGGDVGTHLRNVREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRI					
	orf122ng	DIDARNIDTQIGGDVGTHLRNVRCFEGFLCNHGRIDIDHLPTLRLNALIRRTQKDAAVRI					
		130	140	150	160	170	180
20		190	200	210	220	230	240
	orf122-1.pep	FELCGGVGEMAADIAQTCRTEQRVGNVQQRIGIGVSEQPFKWFDFNSAKYQLSAFQQLV					
	orf122ng	FELCGGVGKMAADVAQTCRTEQRVGNVQQRVGIRMPEQPFKWFDFNSAKYQLSAFQQLV					
25		190	200	210	220	230	240
		250					
	orf122-1.pep	DIVALSDTDVRHRLCSX					
		:					
30	orf122ng	DIVALSDTDIRHRLCSX					
		250					

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### Example 95

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 799>:

35	1	..GCCGGCGCGA	GTGCGAACAA	CATTTCGCG	CGTTTTGCGG	AAACACCCGT
	51	CGCTGTCAGC	GTTACCCTGA	TCGGCAGCGT	ACTTGCCGTC	ATGCTGCCCG
	101	TTACCGAATA	TGAAACTTC	CTGCTGCTTA	TCGGCTCGGT	ATTGCGCCG
	151	ATGGGCGGGA	TTTGATTGC	CGACTTTTC	GTCTGAAAC	GGCGTGA

This corresponds to the amino acid sequence <SEQ ID 800; ORF125>:

40	1	..AGASANNISA	RFAETPVAVS	VTLIGTVLAV	MLPVTEYENF	LLLIGSVFAP
	51	MGGFDCRLFR	LETA*			

Further work revealed the complete nucleotide sequence <SEQ ID 801>:

	1	ATGTCGGGCA	ATGCCTCCTC	TCCTTCATCT	TCCTCCGCCA	TCGGGCTGAT
	51	TTGGTTCGGC	GCGGCGGTAT	CGATTGCCGA	AATCAGCACG	GGTACGCTGC
45	101	TTGCGCCTTT	GGGCTGGCAG	CGCGGTCTGG	CGGCTCTACT	TTTGGGTGAT
	151	GCCGTCGGCG	GCGCGCTGTT	TTTTCGGCGG	GCGTATATCG	GCGCACTGAC
	201	CGGACGCAGC	TCGATGGAAA	GCGTGCGCCT	GTCGTTCCGC	AAACGCGGTT
	251	CAGTGCTGTT	TTCCGTGGCG	AATATGCTGC	AACTGGCCCG	CTGGACGGCG
	301	GTGATGATTT	ACGCGGCGC	AACGGTCAGC	TCCGCTTTGG	GCAAAGTGTT
50	351	GTGGGACGGC	GAATCTTTTG	TCTGGTGGGC	ATTGGCAAAC	GGCGCGCTGA
	401	TTGTGCTGTG	GCTGTTTTTC	GGCGCACGCA	AAACAGGCGG	GCTGAAAACC
	451	GTTTCGATGC	TGCTGATGCT	GTTGGCGGTT	CTGTGGCTGA	GTGCCGAAGT
	501	CTTTTCACG	GACGCGAGCA	CCGCCGCACA	GTTTCAGAC	GGCATGAGTT
	551	TCGGAACGGC	AGTCGAGCTG	TCCGCCGTGA	TGCCGCTTTC	CTGGCTGCCG
55	601	CTTGCCGCGG	ACTACACGCG	CCACGCGCGC	CGCCGTTTTC	CGGCAACCCT
	651	GACGGCAACG	CTCGCTTACA	CGCTGACCGG	CTGCTGGATG	TATGCCTTGG
	701	GTTTGGCAGC	GGCGTTGTTT	ACCGGAGAAA	CCGACGTGGC	AAAAATCCTG
	751	CTGGGCGCAG	GTTTGGGTGC	GGCAGGCATT	TTGGCGGTGC	TCCTCTCCAC

10 This corresponds to the amino acid sequence <SEQ ID 802; ORF125-1>:

15

**Homology with a predicted ORF from *N.meningitidis* (strain A)**

25	orf125.pep				10	20	30
					AGASANNISARFAETPVAVSVTLIGTVLAV		
	orf125a	KILLGAGLGAAGILAVVLSTVTTTFLDAYSAGVSANNISAKLSEIPIAVAVAVVGTLLAV					
		250	260	270	280	290	300
30	orf125.pep		40	50	60		
		MLPVTEYENFLLIGSVFAPMGGFDCRLFRLETAX					
	orf125a	LLPVTEYENFLLIGSVFAPMAAVLIADFFVLKRREEIEG					
		310	320	330	340		

	1	ATGTCGGGCA	ATGCCTCCTC	TCNTTCATCT	TCCGCCGCCA	TCCGGGTGAT
	51	TTGGTTCGGC	GCGGCGGTAT	CGATTGCCGA	AATCAGCACG	GGTACACTGC
	101	TTGCGCCTTT	GGGCTGGCAG	CGCGGTCTGG	CNGCTCTCTG	TTTGGGTCAT
40	151	GCCGTCGGCG	GCGCGCTGTT	TTTTGCGGCG	GCGTATATCG	GCGCACTGAT
	201	CGGACNCANC	TCGATGGAAA	GCGTGC GCCT	GTCGTCTCGC	AAACGCGGTT
	251	CAGTGCTGTT	TTCCGTGGCG	AATATGCTGC	AACTGGCCGG	CTGGACGGCG
	301	GTGATGATTT	ACGCCGGCGC	AACGGTCAGC	TCCGCTTTGG	GCAAAGTGTT
	351	GTGGGACGGC	GAATCTTTTG	TCTGGTGGGC	ATTGGCAAAC	GGCGCGCTGA
45	401	TTGTGCTGTG	GCTGGTTTTT	GGCGCACGCA	AAACAGGCGG	GCTGAAAACC
	451	TTTTTCGATG	TGCTGATGCT	GTTTGGCGTT	CTGTGGCTGA	GTCGCCGAAT
	501	NTTTTCCACG	GACGGCAGCA	CCGCCGCANN	GGTNNCAGAC	GGCATGTAGTT
	551	TCGGAACGGC	AGTCGAGCTG	TCCGCCGTNA	TGCCGCTTTC	TTGGCTGCCG
	601	CTGGCCGCGC	ACTACACGCG	CCACGCGCGC	CGCCCCGTTT	CGGCAACCCT
50	651	GACGGCAACG	CTCGCCTACA	CGCTGACCGG	CTGCTGGATG	TATGCCTTTGG
	701	GTTTGGCAGC	GTCGTTGTTT	ACCGAGAGAA	CCGACGTGGC	AAAAATCCTG
	751	CTGGGCGCAG	GTTTGGGTGC	GGCAGGCATT	TTGGCGGTGC	TCCTGTGCGAC
	801	CGTTACCACC	ACTTTTCTCG	ATGCNACTC	CGCCGGCGTA	AGTGCCAACA
	851	ATATTTCCGC	CAAACCTTCG	GAAATACCNA	TCGCCGTTGC	CGTCGCCGTT
55	901	GTCGGCACAC	TGCTTGCCGT	CCTCTGCCCT	GTTACCGAAT	ATGAAAACTT
	951	CCTGCTGCTT	ATCGGCTCGG	TATTTGCGCC	GATGGCGGCG	GTTTTGATTG
	1001	CCGACTTTTT	CGCTTTGAAA	CGGCGTGAGG	AGATTGAAGG	C..

**This encodes a protein having the partial amino acid sequence <SEQ ID 804>:**

1 MSGNASSXSS SAAIGLIWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH  
51 AVGGALFFAA AYIGALTGXX SMESVRLSFG KRGSVLFSA NMLOLAGWTA



-444-

101 VMYAGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARKTGGLKT  
 151 VSMLLMLLAV LWLSAEXFST AGSTAAXVD GMSEGTAVEL SAVMPLSWLP  
 201 LAADYTRHAR RFFAATLTAT LAYTLTGCM YALGLAAALF TGETDVAKIL  
 251 LGAGLGAAGI LAVVLSTVTT TFLDAYSAGV SANNISAKLS EPIAVAVAV  
 5 301 VGTLLAVLLP VTEYENFLLL IGSVFAPMAA VLIADFFVLK RREEIEG..

ORF125a and ORF125-1 show 94.5% identity in 347 aa overlap:

		10	20	30	40	50	60
10	orf125a.pep	MSGNASSXSSSA	IGLIWFGAAVSIAE	ISTGTLAPL	GWQRLAALLL	GHAVGGALFFAA	
	orf125-1	MSGNASSPSSSA	IGLIWFGAAVSIAE	ISTGTLAPL	GWQRLAALLL	GHAVGGALFFAA	
		10	20	30	40	50	60
15	orf125a.pep	AYIGALTGX	SMESVRLSFG	KRGSVLFS	VANMLQLAG	WTAVMIYAG	TVSSALGKVLWDG
	orf125-1	AYIGALTGR	SSMESVRLSFG	KRGSVLFS	VANMLQLAG	WTAVMIYAG	TVSSALGKVLWDG
		70	80	90	100	110	120
20	orf125a.pep	ESFVWWALAN	GALIVLWLVF	GARKTGGLKT	VSMLLMLLAV	LWLSAEXFST	AGSTAAXVD
	orf125-1	ESFVWWALAN	GALIVLWLVF	GARKTGGLKT	VSMLLMLLAV	LWLSAEVFST	AGSTAQVSD
		130	140	150	160	170	180
25	orf125a.pep	GMSFGTAVEL	SAVMPLSWL	PLAADYTRH	ARRPFAATLT	ATLAYTLTG	CWMYALGLAAALF
	orf125-1	GMSFGTAVEL	SAVMPLSWL	PLAADYTRH	ARRPFAATLT	ATLAYTLTG	CWMYALGLAAALF
		190	200	210	220	230	240
30	orf125a.pep	TGETDVAKIL	LGAGLGAAG	ILAVVLSTV	TTTFLDAYS	SAGVSANNIS	AKLSEIPIAVAVAV
	orf125-1	TGETDVAKIL	LGAGLGAAG	ILAVVLSTV	TTTFLDAYS	SAGASANNIS	ARFAETPVAVGVTL
		250	260	270	280	290	300
35	orf125a.pep	VGTLLAVLLP	VTEYENFLLL	IGSVFAPMA	AVLIADFFVL	KRREEIEG	
	orf125-1	IGTVLAVMLP	VTEYENFLLL	IGSVFAPMA	AVLIADFFVL	KRREEIEG	FDAGLVWLWLAGF
		310	320	330	340	350	360

#### Homology with a predicted ORF from *N.gonorrhoeae*

45 ORF125 shows 86.2% identity over a 65aa overlap with a predicted ORF (ORF125ng) from *N.gonorrhoeae*:

	orf125.pep	AGASANNISARFAETP	VAVSVTLIGTVLAV	30
	orf125ng	KILLGAGLGITGILAV	VLSTVTTTFLDYSAGASANNISARFAEIPVAVGVTL	LIRTVLAV 308
50	orf125.pep	MLPVTYENFLLLIGSV	FAPM-GGFDCRLFRLETA	64
	orf125ng	MLPVTYKNFLLIRSV	FPGMAGGFDCRLFCLKTA	343

55 An ORF125ng nucleotide sequence <SEQ ID 805> was predicted to encode a protein having amino acid sequence <SEQ ID 806>:

	1	MSGNASSPSS	SAAIGLVWFG	AAVSIAEIST	GTLAPLWQ	RGLAALLLGH
	51	AVGGALFFAA	AYIGALTGRS	SMESVRLSFG	KCGSVLFSVA	NMLQLAGWTA
	101	VMYVVGATVS	SALGKVLWDG	ESFVWWALAN	GALIVLWLVF	GARRTGGLKT
	151	VSMLLMLLAV	LWLSVEVEAS	SGTNAAPAVS	DGMTFGTAVE	LSAVMPLSWL
60	201	PLAADYTRQA	RRPFAATLTA	TLAYTLTGCV	MYALGLAAAL	FTGETDVAKI
	251	LLGAGLGITG	ILAVVLSTVT	TTFLDYSAG	ASANNISARF	AEIPVAVGV
	301	LIRTVLAVML	PVTEYKNFLL	LIRSVFGPMA	GGFDCRLFCL	KTA*

Further work revealed the following gonococcal DNA sequence <SEQ ID 807>:

```

      1  ATGTCGGGCA ATGCCTCCTC TCCTTCATCT TCCGCCGCCA TCGGGCTGGT
    51  TTGGTTCGGC GCGGCGGTAT CGATTGCCGA AATCAGCACG GGTACGCTGC
    101  TCGCCCCCTT GGGCTGGCAG CGCGGTCTGG CGGCCCTGCT TTTGGGTGAT
    151  GCCGTCGGCG GCGCGCTGTT TTTTGCGGCG GCGTATATCG GCGCACTGAC
    201  CGGACGCAGC TCGATGGAAA GTGTGCGCCT GTCGTTCCGC AAATGCGGTT
    251  CAGTGCTGTT TTCCGTGGCG AATATGCTGC AACTGGCCGG CTGGACGGCG
    301  GTGATGATTT ACGTCGGCGC AACGGTCAGC TCCGCTTTGG GCAAAGTGTT
    351  GTGGGACGGC GAATCCTTTG TCTGGTGGGC ATTGGCAAAC GGCGCACTGA
    401  TCGTGCTGTG GCTGGTTTTT GCGGCACGCA GAACGGGCGG GCTGAAACC
    451  GTTTCGATGC TGCTGATGCT GCTTGCCGTG TTGTGGTTGA GCGTCGAAGT
    501  GTTCGCTTCG TCCGGCACAA ACGCCGCGCC CGCCGTTTCA GACGGCATGA
    551  CCTTCGGAAC GGCAGTCGAA CTGTCCGCCG TCATGCCGCT TTCCTGGCTG
    601  CCGCTGGCCG CCGACTACAC GCGCCAAGCA CGCCGCCCGT TTGCGGCAAC
    651  CCTGACGGCA ACGCTCGCCT ATACGCTGAC GGGCTGCTGG ATGTATGCCT
    701  TGGGTTTGGC GCGGCTCTG TTTACCGGAG AAACCGACGT GGCGAAATC
    751  CTGTTGGGCG CGGGCTTGGG CATAACGGGC ATTCTGGCAG TCGTCCTCTC
    801  CACCGTTACC ACAACGTTTC TCGATACCTA TTCCGCGGCG GCGAGTGCGA
    851  ACAACATTTC CGCGCGTTTT GCGGAAATAC CCGTCGCTGT CGGCGTTACC
    901  CTGATCGGCA CGGTGCTTGC CGTCATGCTG CCCGTTACCG AATATAAAAA
    951  CTTCTGCTG CTATCGGCT CGGTATTGTC GCCGATGGCG GCGGTTTTGA
   1001  TTTGCGGACT TTTCGCTTTA AAACGGCGTG AGGAGATTGA AGGCTTTGAC
   1051  TTTGCCGAC TGTTCTGTG GCTGGCAGGC TTCATCCTCT ACCGCTTCCT
   1101  GCTCTCGTCC GGTGGGAAA GCAGCATCGG TCTGACCGCC CCCGTAATGT
   1151  CTGCCGTTGC CATTGCCACC GTATCGGTAC GCCTTTTCTT TAAAAAACC
   1201  CAATCTTAC AAAGGAACCC GTCATGA
  
```

This corresponds to the amino acid sequence <SEQ ID 808; ORF125ng-1>:

```

      1  MSGNASSPSS SAAIGLVWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
    51  AVGGALFFAA AYIGALTGRS SMESVRLSFG KCGSVLFSVA NMLQLAGWTA
   101  VMIYVGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARRTGGLKT
   151  VSMLLMLLAV LWLSVEVFAS SGTNAAPAVS DGMTFGTAVE LSAVMPLSWL
   201  PLAADYTRQA RRPFAATLTA TLAYTLTGCV MYALGLAAAL FTGETDVAKI
   251  LLGAGLGITG ILAVVLSTVT TTFLDYSAG ASANNISARF AEIPVAVGVT
   301  LIGTVLAVML PVTEYKNFLL LIGSVFAPMA AVLIADFFVL KRREEIEGFD
   351  FAGLVWLWAG FILYRFLSS GWESSIGLTA PVMSAVAIAT VSVRLFEEKT
   401  QSLQRNPS*
  
```

ORF125ng-1 and ORF125-1 show 95.1% identity in 408 aa overlap:

```

      10      20      30      40      50      60
    40  orf125-1.pep  MSGNASSPSSSSAIGLVWFGAAVVSIAEISTGTLAPLGWQRLAALLLGHAVGGALFFAA
        orf125ng-1  MSGNASSPSSSAAIGLVWFGAAVVSIAEISTGTLAPLGWQRLAALLLGHAVGGALFFAA
      10      20      30      40      50      60

      70      80      90     100     110     120
    45  orf125-1.pep  AYIGALTGRSSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG
        orf125ng-1  AYIGALTGRSSMESVRLSFGKCGSVLFSVANMLQLAGWTAVMIYVGATVSSALGKVLWDG
      70      80      90     100     110     120

      130     140     150     160     170     179
    50  orf125-1.pep  ESFVWWALANGALIVLWLVFGARKTGGLKTVSMLLMLLAVLWLSAEVFSTAGSTAAQ-VS
        orf125ng-1  ESFVWWALANGALIVLWLVFGARRTGGLKTVSMLLMLLAVLWLSVEVFASSGTNAAPAVS
      130     140     150     160     170     180

      180     190     200     210     220     230     239
    55  orf125-1.pep  DGMSFGTAVELSAVMPLSWLPLAADYTRHARRPFAATLTATLAYTLTGCVMYALGLAAAL
        orf125ng-1  DGMTFGTAVELSAVMPLSWLPLAADYTRQARRPFAATLTATLAYTLTGCVMYALGLAAAL
      190     200     210     220     230     240

      240     250     260     270     280     290     299
    60  orf125-1.pep  FTGETDVAKILLGAGLGAAGILAVVLSTVTTTFLDAYSAGASANNISARFAETPVAVGVT
        orf125ng-1  FTGETDVAKILLGAGLGITGILAVVLSTVTTTFLDYSAGASANNISARFAEIPVAVGVT
      250     260     270     280     290     300
  
```

		300	310	320	330	340	350	359
	orf125-1.pep	LIGTVLAVMLPVTEYENFLL	LIGSVFAPMAAVLIADFFVL	KRREEIEGFD	FAGLVLWLAG			
5	orf125ng-1	LIGTVLAVMLPVTEYKFNLL	LIGSVFAPMAAVLIADFFVL	KRREEIEGFD	FAGLVLWLAG			
		310	320	330	340	350	360	
10	orf125-1.pep	FILYRFLSSGWESSIGLTAPVMSAVAIATVSVRLFFKKTQSLQRNPSX						
	orf125ng-1	FILYRFLSSGWESSIGLTAPVMSAVAIATVSVRLFFKKTQSLQRNPSX						
		370	380	390	400			

Based on this analysis, including the presence of putative leader sequence and transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 96

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 809>:

20	1	ATGACCCGTA	TCGCCATCCT	CGGCGGCGGC	CTCTCGGGAA	GGCTGACCGC
	51	GTTGCAGCTT	GCAGAACAAAG	GTTATCAGAT	TGCACTTTTC	GATAAAAGCT
	101	GCCGCGGGG	CGAACACGCC	GCCGCCTATG	TAGCCGCCGC	CATGCTCGCG
	151	CCTGCAGCGG	A.ACGGTCGA	AGCCACGCC	GAAGTGGTCA	GGCTGGGCAG
	201	GCAGAGCATC	CCGCTTTGGC	GCGGCATCCG	ATGCCGTCTG	AACACGCACA
25	251	CGATGATGCA	GGAAAACGGC	AGCCTGATTG	TATGGCACGG	GCAGGACAAG
	301	CCATTATCCA	GCGAGTTCGT	CCGCCATCTC	AAACGCGGCG	GCGT.ACGGA
	351	TGACGAAATC	GTCCGTTGGC	GCGCCGACGA	CATCGCCGAA	CGCGAACCGC
	401	AACTCGGCGG	ACGTTTTTAA	GACGGCATCT	ACCTGCCGAC	CGAAGC.CAG
	451	CTCGACGGGC	GGCAATTATA	GTCTGCACTT	GCCGACGCTT	TGGACGAACT
	501	GAACGTCCCC	TGCCATTGGG	AACACGAATG	CGTCCCCGAA	GCCTGCAAG..

30 This corresponds to the amino acid sequence <SEQ ID 810; ORF126>:

	1	MTRIAILGGG	LSGRLTALQL	AEQGYQIALF	DKSCRRGEHA	AAYVAAAMLA
	51	PAAXTVEATP	EVVRLGRQSI	PLWRGIRCL	NHTMMQENG	SLIVWHQDK
	101	PLSSEFVRHL	KRGGXTDDEI	VRWRADDIAE	REPQLGGRFX	DGIYLPTEXQ
	151	LDGRQLXSAL	ADALDELNVP	CHWEHECVPE	ACK...	

35 Further work revealed the complete nucleotide sequence <SEQ ID 811>:

	1	ATGACCCGTA	TCGCCATCCT	CGGCGGCGGC	CTCTCGGGAA	GGCTGACCGC
	51	GTTGCAGCTT	GCAGAACAAAG	GTTATCAGAT	TGCACTTTTC	GATAAAGGCT
	101	GCCGCGGGG	CGAACACGCC	GCCGCCTATG	TGCGCGCGC	CATGCTCGCG
40	151	CCTGCAGCGG	AAGCGGTCGA	AGCCACGCC	GAAGTGGTCA	GGCTGGGCAG
	201	GCAGAGCATC	CCGCTTTGGC	GCGGCATCCG	ATGCCGTCTG	AACACGCACA
	251	CGATGATGCA	GGAAAACGGC	AGCCTGATTG	TGTGGCACGG	GCAGGACAAG
	301	CCATTATCCA	GCGAGTTCGT	CCGCCATCTC	AAACGCGGCG	GCGTAGCGGA
	351	TGACGAAATC	GTCCGTTGGC	GCGCCGACGA	CATCGCCGAA	CGCGAACCGC
45	401	AACTCGGCGG	ACGTTTTTCA	GACGGCATCT	ACCTGCCGAC	CGAAGGCCAG
	451	CTCGACGGGC	GGCAAATATT	GTCTGCACTT	GCCGACGCTT	TGGACGAACT
	501	GAACGTCCCC	TGCCATTGGG	AACACGAATG	CGTCCCCGAA	GGCCTGCAAG
	551	CCCAATACGA	CTGGCTGATC	GACTGCCGCG	GCTACGGCGC	AAAAACCGCG
	601	TGGAACCAAT	CCCCCGAGCA	CACCAGCACC	CTGCGCGGCA	TACGCGGCGA
50	651	AGTGGCGCGG	GTTTACACAC	CCGAAATCAC	GCTCAACCGC	CCCGTGCGTC
	701	TGCTCCATCC	GCGTTATCCG	CTCTACATCG	CCCCGAAAGA	AAACCACGTC
	751	TTGTCATCG	GCGCGACCCA	AATCGAAAGC	GAAAGCCAAG	CCCCGCCAG
	801	CCTGCGTTCA	GGGTTGGAAC	TCTTGTCGCG	ACTCTATGCC	ATCCACCCCG
	851	CCTTCGGCGA	AGCCGACATC	CTCGAAATCG	CCACCGGCCT	GCGCCCCACG
	901	CTCAACCAAC	ACAACCCCGA	AATCCGTTAC	AACCGCGCCC	GACGCCTGAT
55	951	TGAAATCAAC	GGCCTTTTCC	GCCACGGTTT	CATGATCTCC	CCCGCCGTAA
	1001	CCGCGCGCGC	CGCCAGATTG	GCAGTGGCAC	TGTTTGACGG	AAAAGACGCG
	1051	CCCGAACGCG	ATAAAGAAAG	CGGTTTGGCG	TATATCCGAA	GACAAGATTA
	1101	A				

This corresponds to the amino acid sequence <SEQ ID 812; ORF126-1>:

```

1  MTRIAILGGG LSGRLTALQL AEQGYQIALF DKGCRERGEHA AAYVAAAML
51 PAEEAVEATP EVVRLGRQSI PLWRGIRCL NTHMTMQENG SLIVWHGQDK
101 PLSSEFVRHL KRGGVADDEI VRWRADDIAE REPQLGGRFS DGIYLPTEGQ
151 LDGRQILSAL ADALDELNVP CHWEHECVPE GLQAQYDWLI DCRGYGAKTA
201 WNQSPEHTST LRGIRGEVAR VYTPEITLNR PVRLHPRYP LYIAPKENHV
251 FVIGATQIES ESQAPASVRS GLELLSALYA IHPAFGEADI LEIATGLRPT
301 LNHHPFIRY NRARRLIEIN GLFRHGFMS PAVTAAARL AVALFDGKDA
351 PERDKESGLA YIRROD*

```

10 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF126 shows 90.0% identity over a 180aa overlap with an ORF (ORF126a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
15	orf126.pep	MTRIAILGGG	LSGRLTALQL	AEQGYQIALF	DKGCRERGEHA	AAYVAAAML	LAPAAXTVEATP
	orf126a						
		10	20	30	40	50	60
20	orf126.pep	EVVRLGRQSI	PLWRGIRCLN	NTHMTMQENG	SLIVWHGQDK	PLSSEFVRHL	KRGGXTDDEI
	orf126a	EVVRLGRQXI	PLWRGIRCHL	KTPAMMXENG	SLIVWHGQDK	PLSNEFVRHL	KRGGVADDDXI
		70	80	90	100	110	120
25	orf126.pep	VRWRADDIAE	REPQLGGRFX	DGIYLPTEQ	LDGRQLXSAL	ADALDELNVP	CHWEHECVPE
	orf126a	VRWRADDIAE	REPQLGGRFSD	GDIYLPTEG	LDGRQILSAL	ADALDELNVP	CHWEHECAPE
		130	140	150	160	170	180

The complete length ORF126a nucleotide sequence <SEQ ID 813> is:

```

1  ATGACCCGTA TCGCCATCCT CGGCGGCGGC CTCTCNGGAA GGCTGACCGC
51 ACTGCAGCTT GCAGAACAAG GTTATCAGAT TGCACCTTTC GATAAAGGCT
101 GCCGCCGGGG CGAACACGCC GCCGCCATG TTGCCCGCG CATGCTCGCG
151 CCTGCGCGGG AAGCGGTCTGA AGCCACGCCT GAAGTGGTCA GGCTGGGCAG
201 GCAGANATC CCGCTTTGGC GCGGCATCCG ATGCCATCTG AAAACGCCTG
251 CCATGATGCA NGAACACGGC AGCCTGATTG TGTGGCACGG GCAGGACAAA
301 CCTTTATCCA ACGAGTTCTG CCGCCATCTC AAACGCGGCG GCGTAGCGGA
351 TGACNAAATC GTCCGTGGC GCGCCGACGA CATCGCCGAA CGCGAACCAG
401 AACTCGCGGG ACGTTTTTCA GACGGCATCT ACCTGCCGAC CGAAGGCCAG
451 CTCGACGGGC GGCAATATT GTCTGCACTT GCCGACGCTT TGGACGAAT
501 GAACGTCCCC TGCCATTGGG AACACGAATG TGCCCCGAA GACTTGCAAG
551 CCCAATACGA CTGGCTGATC GACTGCCGCG GCTACGGCGC AAAAACCGCG
601 TGGAACCAAT CCCCAGANNA NACCAGCACC CTGCGCGGCA TACGCGGCGA
45 651 AGTGGCGCGG GTTTACACAC CCGAAATCAC GCTCAACCGC CCCGTGCGCC
701 TGCTACACCC GCGCTATCCG CTNTACATCG CCCCAGAAAG AAACNCGTG
751 TTCGTCATCG GCGCGACCCA AATCGAAAGC GAAAGCCAAG CACCTGCCAG
801 CGTGCGTTCC GGGCTGGAAC TCTTATCCGC ACTCTATGCC GTCCACCCCG
851 CCTTCGGCGA AGCCGACATC CTCGAAATCG CCACCGGCCT GCGCCCCACG
50 901 CTCAATCACC ACAACCCCGA AATCCGTTAC AACCGCGCCC GACGCCTGAT
951 TGAAATCAAC GGCCTTTTCC GCCACGGTTT CATGATCTCC CCGCCGTAA
1001 CCGCCGCCGC CGTCAGATTG GCAGTGGCAC TGTGTGACGG AAAAGANGCG
1051 CCCGAACGCG ATGAAGAAAG CGGTTTGCG TATATCCGAA GACAAGATTA
1101 A

```

55 This encodes a protein having amino acid sequence <SEQ ID 814>:

```

1  MTRIAILGGG LSGRLTALQL AEQGYQIALF DKGCRERGEHA AAYVAAAML
51 PAEEAVEATP EVVRLGRQXI PLWRGIRCHL KTPAMMXENG SLIVWHGQDK
101 PLSNEFVRHL KRGGVADDDXI VRWRADDIAE REPQLGGRFS DGIYLPTEGQ
151 LDGRQILSAL ADALDELNVP CHWEHECAPE DLQAQYDWLI DCRGYGAKTA
60 201 WNQSPXXTST LRGIRGEVAR VYTPEITLNR PVRLHPRYP LYIAPKENXV

```

251 FVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIATGLRPT  
 301 LNHHNPEIRY NRARRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKXA  
 351 PERDEESGLA YIRRQD\*

ORF126a and ORF126-1 show 95.4% identity in 366 aa overlap:

5		10	20	30	40	50	60
	orf126a.pep	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAEAVEATP					
	orf126-1	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAEAVEATP					
10		10	20	30	40	50	60
	orf126a.pep	70	80	90	100	110	120
	orf126-1	EVVRLGRQXIPLWRGIRCHLKTAMMXENGSLIVWHGQDKPLSNEFVRHLKRGGVADDDXI					
15		70	80	90	100	110	120
	orf126a.pep	EVVRLGRQSIPLWRGIRCLNTHMTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI					
	orf126-1	EVVRLGRQSIPLWRGIRCLNTHMTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI					
20		130	140	150	160	170	180
	orf126a.pep	VRWRADDIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPE					
	orf126-1	VRWRADDIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECVPE					
25		190	200	210	220	230	240
	orf126a.pep	DLQAQYDWLIDCRGYGAKTAWNQSPXSTLRLGIRGEVARVYTPEITLNRPVRLHPRYP					
	orf126-1	GLQAQYDWLIDCRGYGAKTAWNQSPHTSTLRLGIRGEVARVYTPEITLNRPVRLHPRYP					
30		250	260	270	280	290	300
	orf126a.pep	LYIAPKENXVVFVIGATQIESESQAPASVRSGLLELLSALYAVHPAFGEADILEIATGLRPT					
	orf126-1	LYIAPKENHVVFVIGATQIESESQAPASVRSGLLELLSALYAIHPAFGEADILEIATGLRPT					
35		310	320	330	340	350	360
	orf126a.pep	LNHHNPEIRYNRARRLIEINGLFRHGFMSIPAVTAAAVRLAVALFDGKXAPERDEESGLA					
	orf126-1	LNHHNPEIRYNRARRLIEINGLFRHGFMSIPAVTAAARLAVALFDGKDAPERDKESGLA					
40		310	320	330	340	350	360
	orf126a.pep	YIRRQDX					
	orf126-1	YIRRQDX					

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF126 shows 90% identity over a 180 aa overlap with a predicted ORF (ORF126ng) from *N.gonorrhoeae*:

50	orf126.pep	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAAYVAAAMLAPAAXTVEATP	60
	orf126ng	MTRIAVLGGGLSGRLTALQLAEQGYQIELFDKGTROGEHAAAYVAAAMLAPAAEAVEATP	60
55	orf126.pep	EVVRLGRQSIPLWRGIRCLNTHMTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGXTDDEI	120
	orf126ng	EVIRLGRQSIPLWRGIRCLNTLTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI	120
60	orf126.pep	VRWRADDIAEREPQLGGRFXDGIYLPTEQQLDGRQLXSALADALDELNVPCHWEHECVPE	180
	orf126ng	VRWRADEIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPO	180

An ORF126ng nucleotide sequence <SEQ ID 815> was predicted to encode a protein having amino acid sequence <SEQ ID 816>:

1 MTRIAVLGGG LSGRLTALQL AEQGYQIELF DKGTROGEHA AAYVAAAMLA

51 PAAEAVEATP EVIRLGRQSI PLWRGIRCL NTLTMMQENG SLIVWHGQDK  
 101 PLSSEFVRHL KRGGVADDEI VRWRADEIAE REPQLGGRFS DGIYLPTEGQ  
 151 LDGRQILSAL ADALDELNVP CHWEHECAPQ DLQAQYDWVI DCRGYGAKTA  
 201 WNQSPEHTST LRIGIRGEVRG FTRPKSRSTA PCACCTRAIR STSPRKKTTS  
 251 SSSARPKSKA KAKPPPAYVP GWNSYPRSMPT STPPSAKPTS SKWRPGLRPT  
 301 LNHHNPEIRY SRERRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKDA  
 351 PERDEESGLA YIGRQD\*

Further work revealed the following gonococcal DNA sequence <SEQ ID 817>:

10 1 ATGACCCGTA TCGCCGTCCT CGGAGGCGGC CTTTCCGGAA GGCTGACCGC  
 51 ATTGCAGCTT GCAGAACAAG GTTATCAGAT TGAACCTTTC GACAAGGGCA  
 101 CCCGCCAAGG CGAACACGCC GCCGCCTATG TTGCCGCCGC GATGCTCGCG  
 151 CCTGCGGCGG AAGCGGTGCA GGCAACGCCG GAAGTCATCA GGCTGGGCAG  
 201 GCAGAGCATT CCGCTTTGGC GCGGCATCCG ATGCCGTCTG AACACGCTCA  
 251 CGATGATGCA GGAAAACGGC AGCCTGATG TGTGGCACGG GCAGGACAAG  
 15 301 CCATTATCCA GCGAGTTCGT CCGCCATCTC AAACGCGGCG GCGTAGCGGA  
 351 TGACGAAATC GTCCGTTGGC GCGCCGATGA AATCGCCGAA CGCGAACCGC  
 401 AACTCGGCGG ACCTTTTTC GACGGCATCT ACCTGCCGAC CGAAGGCCAG  
 451 CTCGACGGGC GGCAATATT GTCTGCACTT GCCGACGCTT TGGACGAACT  
 501 GAACGTCCCT TGCCATTGGG AACACGAATG CGCCCCCAA GACCTGCAAG  
 20 551 CCCAATACGA CTGGGTAAATC GACTGCCGGG GCTACGGGCG GAAAACCGCG  
 601 TGGAAACCAAT CCCCCGAGCA CACCAGCACC TTGCGCGGCA TACGCGGCCA  
 651 AGTGGCGCGG GTTTACACGC CCGAAATCAC GCTCAACCGC CCCGTGCGCC  
 701 TGCTGCACCC GCGCTATCCG CTCTACATCG CCCCAGAAAG AAACCACGTC  
 751 TTCGTCATCG GCGCGACCCA AATCGAAAGC GAAAGCCAAG CCCCCGCCAG  
 25 801 CGTACGTTCC GGGCTGGAAC TCTTATCCG GCTCTATGCC GTCCACCCCG  
 851 CCTTCGCGCA AGCCGACATC CTCGAAATCG CCGCCGGCCT GCGCCCCACG  
 901 CTCAACCACC ACAACCCGCA AATCGGCTAC AGCCGCGAAC GCCGCCTCAT  
 951 CGAAATCAAC GGCCTTTTCC GGCACGGCTT TATGATTTC CCGCCGTAA  
 1001 CCGCCGCCGC CGTCAGATTG GCAGTGGCAC TGTTTGACGG AAAAGACGCG  
 30 1051 CCCGAACGTG ATGAAGAAAG CGGTTTGGCG TATATCGGAA GACAAGATTA  
 1101 A

This corresponds to the amino acid sequence <SEQ ID 818; ORF126ng-1>:

1 MTRIAVLGGG LSGRLTALQL AEQGYQIELF DKGTRQGEHA AAYVAAAMLA  
 51 PAAEAVEATP EVIRLGRQSI PLWRGIRCL NTLTMMQENG SLIVWHGQDK  
 35 101 PLSSEFVRHL KRGGVADDEI VRWRADEIAE REPQLGGRFS DGIYLPTEGQ  
 151 LDGRQILSAL ADALDELNVP CHWEHECAPQ DLQAQYDWVI DCRGYGAKTA  
 201 WNQSPEHTST LRIGIRGEVAR VYTPEITLNR PVRLHPRYP LYIAPKENHV  
 251 FVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIAAGLRPT  
 301 LNHHNPEIRY SRERRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKDA  
 40 351 PERDEESGLA YIGRQD\*

ORF126ng-1 and ORF126-1 show 95.1% identity in 366 aa overlap:

		10	20	30	40	50	60
orf126-1.pep		MTRIAVLGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAAEAVEATP					
45 orf126ng-1		MTRIAVLGGGLSGRLTALQLAEQGYQIELFDKGTQGEHAAAYVAAAMLAPAAEAVEATP					
		10	20	30	40	50	60
		70	80	90	100	110	120
orf126-1.pep		EVVRLGRQSIPLWRGIRCLNTHNTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI					
50 orf126ng-1		EVIRLGRQSIPLWRGIRCLNTLTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI					
		70	80	90	100	110	120
		130	140	150	160	170	180
orf126-1.pep		VRWRADDIAEREPLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECVPE					
55 orf126ng-1		VRWRADEIAEREPLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPQ					
		130	140	150	160	170	180
		190	200	210	220	230	240
orf126-1.pep		GLQAQYDWLIDCRGYGAKTAWNQSPEHTSTLRIGIRGEVARVYTPEITLNRPVRLHPRYP					
60 orf126ng-1		DLQAQYDWVIDCRGYGAKTAWNQSPEHTSTLRIGIRGEVARVYTPEITLNRPVRLHPRYP					
		190	200	210	220	230	240

-450-

		250	260	270	280	290	300
	orf126-1.pep	LYIAPKENHVFVIGATQIESESQAPASVRSGLELLSALYAIHPAFGEADILEIATGLRPT					
5	orf126ng-1	LYIAPKENHVFVIGATQIESESQAPASVRSGLELLSALYAVHPAFGEADILEIAAGLRPT					
		250	260	270	280	290	300
	orf126-1.pep	LNHHNPEIRYNRRARLIEINGLFRHGFMI SPAVTAAARLAVALFDGKDAPERDKESGLA					
10	orf126ng-1	LNHHNPEIRYSRERRLIEINGLFRHGFMI SPAVTAAVRLAVALFDGKDAPERDEESGLA					
		310	320	330	340	350	360
15	orf126-1.pep	YIRRQDX					
	orf126ng-1	YIGRQDX					

Furthermore, ORF126ng-1 shows homology to a putative *Rhizobium* oxidase flavoprotein:

	gi 2627327 (AF004408) putative amino acid oxidase flavoprotein [Rhizobium etli]
	Length = 327
20	Score = 169 bits (423), Expect = 3e-41
	Identities = 112/329 (34%), Positives = 163/329 (49%), Gaps = 25/329 (7%)
	Query: 3 RIAVLGGGLSGRLTALQLAEQGYQIELFDKGTROGEHXXXXXXXXXXXXXXXXXXXXX 62
	RI V G G++G A QL G+++ L ++ G
25	Sbjct: 2 RILVNGAGVAGLTVAWQLYRHGFRVTLAERAGTVGA-GASGFAGGMLAPWCERESAEEPV 60
	Query: 63 IRLGRQSIPLWRGIRCLNLTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEIVR 122
	+ LGR + W + G+L+V G+D F R G DE+
30	Sbjct: 61 LTLGRLAADWWEAA-----LPGHVHRRGTLVVGAGGRDTGELDRFSRRTS-GWEWLDEVA- 113
	Query: 123 WRADEIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPQDL 182
	IA EP L GRF ++ E LD RQ L+ALA L++ +
	Sbjct: 114 -----IAALEPDLAGRFRRLFFRQEAHLDPQALAAALAGLEDARMRLTLG---VVGES 165
35	Query: 183 QAQYDWVIDCRGYGAKTAWNQSPEHTSTLRGIRGEVARVYTPFITLNRPVRLHLHPRYPLY 242
	+D V+DC G LRG+RGE+ V T E++L+RPVRLHLHPR+P+Y
	Sbjct: 166 DVDHDRVVDCTGAA-----QIGRLPGLRGVGEMLCVETTEVSLSRPVRLHLHPRHIY 218
40	Query: 243 IAPKENHVFVIGATQIESESQAPASVRSGLELLSALYAVHPAFGEADILEIAAGLRPTLN 302
	I P++ + F++GAT IES+ P + RS +ELL+A YA+HPAFGEA + E AG+RP
	Sbjct: 219 IVPDRKNRFMVGATMIESDDGGPITARSIMELLNAAAYAMHPAFGEARVTETGAGVRPAYP 278
	Query: 303 HHNPEIRYSRERRLIEINGLFRHGFMI SP 331
	+ P R ++E R + +NGL+RHGF+++P
45	Sbjct: 279 DNLP--RVTQEGRTLHVNGLYRHGFLAP 305

This analysis suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

### Example 97

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID

50	819>:	
	1	ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
	51	GATATTGTCT GTACTTGCTT TAATTGTTA TCCGAGCTAT CGCAATTATG
	101	TTGAGAAAGC AAAGATAAAT GCAGTGCGGG CAGCCTTGT AGAAAATGCA
55	151	CATTTTATGG AAAAGTTTTA TCTGCAGAA GGGAGGTTTA AACAAACATC
	201	TACCAAGTGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
	251	GTTTGAATGG AATCGtCGCG CGGG..GCTT TAGACAGTAA ATTCATGTTG
	301	AAGCGGCTAG CCATAGATAA AGATAAAAT CCTTTTATTA TTAAGATGAA
	351	TGAAAATCTA GTAACCTTTA aTTTGCAAGA AGTCCGCCAG TTCGTGTAGT
	401	GACGGGCTGG ATTATTTTAA AGGAAATGAT AAGGACTGCA AGTTACTTAA
60	451	GTAG

This corresponds to the amino acid sequence <SEQ ID 820; ORF127>:

```

1 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLEN
51 HFMEKFYLN GRFKQTSTKW PSLPIKEAEG FCIRLNGIVA RXALDSKFML
101 KAVAIDKDKN PFIKMENLV VTFICKKSAS SCSDGLDYFK GNDKCKLLK
151 *

```

Further work revealed the following DNA sequence <SEQ ID 821>:

```

1 ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
51 GATATTGTCT GACTTGCTT TAATTGTTA TCCGAGCTAT CGCAATTATG
101 TTGAGAAAGC AAAGATAAAT GCAGTGC GAGCCTTGT AGAAAATGCA
151 CATTTTATGG AAAAGTTT TCTGCAGAAAT GGGAGGTTT AACAAACATC
201 TACCAAGTGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
251 GTTTGAATGG AATCGCGCGC GGGGCTTTAG ACAGTAAATT CATGTTGAAG
301 GCGGTAGCCA TAGATAAAGA TAAAATCCT TTTATTATTA AGATGAATGA
351 AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAAGTACG
401 GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG

```

This corresponds to the amino acid sequence <SEQ ID 822; ORF127-1>:

```

1 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLEN
51 HFMEKFYLN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
101 AVAIDKDKNP FIKMENLV TFICKKSASS CSDGLDYFKG NDKCKLLK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF127 shows 98.0% identity over a 150aa overlap with an ORF (ORF127a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
25	orf127.pep	MTDNRGFTLV	ELISVVLILS	VLALIVYPSY	RNYVEKAKIN	AVRAALLEN	HFMEKFYLN
	orf127a	MTDNRGFTLV	ELISVVLILS	VLALIVYPSY	RNYVEKAKIN	TVRAALLEN	HFMEKFYLN
		10	20	30	40	50	60
30	orf127.pep	GRFKQTSTKW	PSLPIKEAEG	FCIRLNGIV	ARXALDSKFML	KAVAIDKDKN	PFIKMENLV
	orf127a	GRFKQTSTKW	PSLPIKEAEG	FCIRLNGI	-ARGALDSKFML	KAVAIDKDKN	PFIKMENLV
		70	80	90	100	110	120
35	orf127.pep	VTFICKKSASS	CDGLDYFKG	NDKCKLLKX			
	orf127a	VTFICKKSASS	CDGLDYFKG	NDKCKLLKX			
		130	140	150			
40		120	130	140	150		

The complete length ORF127a nucleotide sequence <SEQ ID 823> is:

```

1 ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
51 GATATTGTCT GACTTGCTT TAATTGTTA TCCGAGCTAT CGCAATTATG
101 TTGAGAAAGC AAAGATAAAT ACAGTGC GAGCCTTGT AGAAAATGCA
151 CATTTTATGG AAAAGTTT TCTGCAGAAAT GGGAGGTTT AACAAACATC
201 TACCAAAATGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGTATCC
251 GTTTGAATGG AATCGCGCGC GGGGCTTTAG ACAGTAAATT CATGTTGAAG
301 GCGGTAGCCA TAGATAAAGA TAAAATCCT TTTATTATTA AGATGAATGA
351 AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAAGTACG
401 GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG

```

This encodes a protein having amino acid sequence <SEQ ID 824>:

```

1 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN TVRAALLEN
51 HFMEKFYLN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
101 AVAIDKDKNP FIKMENLV TFICKKSASS CSDGLDYFKG NDKCKLLK*

```



-452-

ORF127a and ORF127-1 show 99.3% identity in 149 aa overlap:

		10	20	30	40	50	60
5	orf127a.pep	MTDNRGFTLVELISVVLILSVLALIVYPSYRNYVEKAKINTVRAALLENNAHFMEKFYLQN					
	orf127-1	MTDNRGFTLVELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLENNAHFMEKFYLQN					
		10	20	30	40	50	60
10	orf127a.pep	GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFIKMNENLV					
	orf127-1	GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFIKMNENLV					
		70	80	90	100	110	120
15	orf127a.pep	TFICKKSASSCSDGLDYFKGNDKDKCKLLKX					
	orf127-1	TFICKKSASSCSDGLDYFKGNDKDKCKLLKX					
		130	140	150			

## 20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF127 shows 97.3% identity over a 150 aa overlap with a predicted ORF (ORF127ng) from *N.gonorrhoeae*:

	orf127.pep	MTDNRGFTLVELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLENNAHFMEKFYLQN	60
25	orf127ng	MTDNRGFTLVELISVVLILSVLALIVYPSYRNYVEKAKINAVRAAFLENAHFMEKFYLQN	60
	orf127.pep	GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFIKMNENL	120
30	orf127ng	GRFKQTSTKWPSLPIKEAEGFCIRLNGI-ARGALDSKFMLKAVAIDKDKNPFIKMNENL	119
	orf127.pep	VTFICKKSASSCSDGLDYFKGNDKDKCKLLK	150
	orf127ng	VTFICKKSASSCSDRLDYFKGNDKDKCKLLK	149

The complete length ORF127ng nucleotide sequence <SEQ ID 825> is:

35	1	ATGACTGATA	ATCGGGGGTT	TACTGTGTT	GAATTAATAT	CAGTGGTCTT
	51	GATATTGTCT	GTACTTGTCT	TAATTGTTA	TCCGAGCTAT	CGCAATTATG
	101	TTGAGAAAGC	AAAGATAAAT	GCAGTGGGG	CAGCCTTGT	AGAAAATGCA
	151	CATTTTATGG	AAAAGTTTA	TCTGCAGAA	GGGAGATTTA	AACAAACATC
40	201	TACCAAATGG	CCAAGTTTGC	CGATTAAAGA	GGCAGAAGGC	TTTTGTATCC
	251	GTTTGAATGG	AATCGCGCGC	GGGGCTTTAG	ACAGTAAATT	CATGTTGAAG
	301	GCGGTAGCCA	TAGATAAAGA	TAAAAATCCT	TTTATTATTA	AGATGAATGA
	351	AAATCTAGTA	ACCTTTATTT	GCAAGAAGTC	CGCCAGTTCG	TGTAGTGACG
	401	GGCTGATTA	TTTTAAAGGA	AATGATAAGG	ACTGCAAGTT	ACTTAAGTAG

This encodes a protein having amino acid sequence <SEQ ID 826>:

45	1	MTDNRGFTLV	ELISVVLILS	VLALIVYPSY	RNYVEKAKIN	AVRAAFLENA
	51	HFMEKFYLQN	GRFKQTSTKW	PSLPIKEAEG	FCIRLNGIAR	GALDSKFMLK
	101	AVAIDKDKNP	FIKMNENLV	TFICKKSASS	CSDRLDYFKG	NDKDKCKLLK*

ORF127ng and ORF127-1 show 100.0% identity in 149 aa overlap:

50	orf127-1.pep	MTDNRGFTLVELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLENNAHFMEKFYLQN					
	orf127ng-1	MTDNRGFTLVELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLENNAHFMEKFYLQN					
		10	20	30	40	50	60
55	orf127-1.pep	GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFIKMNENLV					
	orf127ng-1	GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFIKMNENLV					
		70	80	90	100	110	120
60							

-453-

```

              130      140      150
orf127-1.pep  TFICKKSASSCSDGLDYFKGNDKDKCLLKX
              |||||
orf127ng-1    TFICKKSASSCSDGLDYFKGNDKDKCLLKX
              130      140      150

```

This analysis, including the fact that the predicted transmembrane domain is shared by the meningococcal and gonococcal proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 10 Example 98

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 827>

```

1  ..GTGTCGCTGG CTTGGGTGAT TGCCTCTCAA ATCTTCCTTT ACGAAGATTT
51  CAACCAAATG CGGAAAACCG GTGGAGCTAT CTGCGGTTTT CTTGTCCAAT
101 ATTTATCTGG GGTTCAGCA GGGGTATTTC GATTGAGTG CCGACGAGAA
151 CCCCGTACTG CATATCTGGT CTTTGGCAGT AGAGGAACAG TATTACCTCC
201 TGATCCCCCT TTTGCTGATA TTTTGCTGCA AAAAAACCAA ATCGCTACGG
251 GTGCTGCGTA ACATCAGCAT CATCCTGTTT TTGATTTTGA CTGCCTCATC
301 GTTTTTGCCA AGCGGGTTTT ATACCGACAT CCTCAACCAA CCAATACTT
351 ATTACCTTTC GACACTGAGG TTTCCCGAGC TGTGGCAGG TTCGCTGCTG
401 GCGGTTTACG GGCAACGCA AAACGGCAGA CGGCAACAG CAAATGGAAA
451 ACGGCAGTTG CTTTCATCAC TCTGCTTCGG CGCATTGCTT GCCTGCCTGT
501 TCGTGATTGA CAAACACAAT CCGTTTATCC CGGGAATGAC CCTGCTCCTT
551 CCCTGCCTGC TGACGGCACT GCTTATCCGG AGTATGCAAT ACGGGACACT
601 TCCGACCCGC ATCCTGTCGG CAAGCCCCAT CGTATTGTGC GGCAAAATCT
651 CTTATTCCCT ATACCTGTAC CATPGGATT TTATTGCTTT CGCTCCGCTC
701 ATTAGAGCGG GGAAACAGCT CGGACTGCCT GCCG..

```

This corresponds to the amino acid sequence <SEQ ID 828; ORF128>:

```

1  ..VSLASVIASQ IFLYEDFNQM RKTVELSAVF LSNIYLGFOQ GYFDLSADEN
51  PVLHIWSLAV EEQYLLYPL LLIFCCKKTK SLRVLRNISI ILFLILTASS
101 FLPSGFYTDI LNQPNTYYLS TLRFPPELLAG SLLAVYQOTQ NGRRQTANGK
151 RQLLSLCFG ALLACLFLVID KHNPFIPGMT LLLPCLLTAL LIRSMQYGTL
201 PTRILSASPI VEVGKISYSL YLYHWIFIAF APLIRGGKQL GLPA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 829>:

```

1  ATGCAAGCTG TCCGATACAG ACCGGAATTT GACGGATTGC GGGCCGTCGC
35  51  CGTGCTATCC GTCATGATTT TCCACCTGAA TAACCGCTGG CTGCCCGGAG
101  GATTCCTGGG GGTGGACATT TTCTTTGTCA TCTCAGGATT CCTCATTACC
151  GGCATCATTG TTTCTGAAAT ACAGAACGGT TCTTTTCTT TCCGGGATTT
201  TTATACCCGC AGGATTAAGC GGATTTATCC TGCCTTTATT GCGGCCGTGT
251  CGCTGGCTTC GGTGATTGCC TCTCAAATCT TCCTTACGA AGATTTCAAC
40  301  CAAATGCGGA AAACCGTGGA GCTTCTGCG GTTTTCTTGT CCAATATTTA
351  TCTGGGGTTT CAGCAGGGGT ATTTGATTT GAGTGCCGAC GAGAACCCCG
401  TACTGCATAT CTGGTCTTTG GCAGTAGAGG AACAGTATTA CCTCCTGTAT
451  CCCCTTTTGC TGATATTTTG CTGCAAAAAA ACCAAATCGC TACGGGTGCT
501  GCGTAACATC AGCATCATCC TGTTTTGTAT TTTGACTGCC TCATCGTTTT
45  551  TGCCAAGCGG GTTTTATACC GACATCCTCA ACCAACCCTA TACTTATTAC
601  CTTTCGACAC TGAGGTTTCC CGAGCTGTTG GCAGGTTTCG TGCTGGCGGT
651  TTACGGGCAA ACGCAAAACG GCAGACGGCA AACAGCAAT GGAAAACGGC
701  AGTTGCTTTC ATCACTCTGC TTCGGCGCAT TGCTTGCCTG CCTGTTCTGT
751  ATTGACAAAC ACAATCCGTT TATCCCGGGA ATGACCCTGC TCCTTCCCTG
50  801  CCTGCTGACG GCACTGCTTA TCCGGAGTAT GCAATACGGG ACACCTCCGA
851  CCCGCATCCT GTCGGCAAGC CCCATCGTAT TTGTCGGCAA AATCTCTTAT
901  TCCCTATACC TGTACCATTG GATTTTATT GCTTTCGCCC ATTACATTAC
951  AGGCGACAAA CAGCTCGGAC TGCCGCGGTT ATCGGCGGTT GCCGCGTTGA
1001 AGGCCGGATT TTCCTGTTG AGTTATTATT TGATTGAACA GCCGCTTAGA
55  1051 AAACGGAAGA TGACCTTCAA AAAGGCATTT TTCTGCCTCT ATCTCGCCCC
1101 GTCCCTGATA CTTGTCGGTT ACAACCTGTA CGCAAGGGGG ATATTGAAC
1151 AGGAACACCT CCGCCCGTTG CCCGGCGCGC CCCTTGCTGC GGAAATCAT

```

15 This corresponds to the amino acid sequence <SEQ ID 830; ORF128-1>:

20

25

Computer analysis of this amino acid sequence gave the following results:

30 Homology with hypothetical integral membrane protein HI0392 of *H.influenzae* (accession number U32723)

**ORF128 and HI0392 show 52% aa identity in 180aa overlap:**

35

Orf128:	1	VSLASVIASQIFLYEDFNQMRKTVLSAVFLSNIYLGQQGYFDLSADENPVLHIWLSLAV	60
		++L S IAS IF+Y DFN++RKT+EL+ FLSN YLG QGYFDLSA+ENPVLHIWLSLAV	
HI0392:	46	MALVSFTIASAIFIYDNDFNKLRTKTIELAI AFLSNFYLGLTQGYFDLSANENPVLHIWLSLAV	105
Orf128:	61	EEQXXXXXXXXXIFCCKTKTSLRVLNRNISIILFLILTASSFLPSGFTDILNQPNNTYYLS	120
		E Q I KK + ++VL I++ILF IL A+SF+ + FY ++L+QPN YYLS	
HI0392:	106	EGQYYLIFPLILILAYKKFREVKVLFITITLILFFILLATSFVSANFYKEVLHQPNIIYYLS	165
Orf128:	121	TLRFPELLAGSLLAVYGQTQNGRRQTANGKRQLSSSLCFGALLACLFVIDKHNPFIPGMT	180
		LRFPPELL GSLLA+Y N + Q + +L+ L L +CLF+++ + FIPG+T	
HI0392:	166	NLRFPPELLVGSLLAIYHNLSN-KVQLSKQVNNILAILSTLLLFSCFLMNNNIAIFIPGIT	224

40

**Homology with a predicted ORF from *N.meningitidis* (strain A)**

45 ORF128 shows 98.0% identity over a 244aa overlap with an ORF (ORF128a) from strain A of *N.*

**meningitidis:**

50

```

                                10      20      30
orf128.pep                      VSLASVIASQIFLYEDFNQMRKTVELSAVF
                                |||||
orf128a      ILSEIQNGSFSFRDFYTRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVF
                60      70      80      90      100     110

```

55

```

              40          50          60          70          80          90
orf128.pep  LSNIYLGFGQGYFDLSADENPVLHIWLSLAVEEQYLLYPILLIFCCKKTKSLRVLRNISI
              |||||
orf128a     LSNIYLGFGQGYFDLSADENPVLHIWLSLAVEEQYLLYPILLIFCCKKTKSLRVLRNISI
              120        130        140        150        160        170

```

60

```

      100      110      120      130      140      150
orf128.pep  ILFLILTASSFLPSGFYTDILNQPNITYYLSLRFPELLAGSLLAVYGQTQNGRRQTANGK
            |||||:|||||

```

-455-

5	orf128a	ILFLILTATSF	LP	SGFYTDIL	NQ	PNTYYL	STL	RFPELLAGS	LL	AVYGQTQ	NGRRQTANGK
		180	190	200	210	220	230				
10	orf128.pep	RQLLSSLCFG	GALLAC	LFVIDKH	NPFIPG	MTLLLPCL	L	TALLIRSM	QYGTLP	TRILSASPI	
		160	170	180	190	200	210				
15	orf128a	RQLLSSLCFG	GALLAC	LFVIDKH	NPFIPG	MTLLLPCL	L	TALLIRSM	QYGTLP	TRILSASPI	
		240	250	260	270	280	290				
20	orf128.pep	VFVGKISYS	LYLHWIF	IAFAPIR	G	GKQLGLPA					
		220	230	240							
25	orf128a	VFVGKISYS	LYLHWIF	IAFAHYIT	GDKQLGL	PAVSAVA	AALTAGF	SLLSYLIE	QPLRKR		
		300	310	320	330	340	350				
30	orf128a	KMTFKKAFF	CLYLAP	SLLIVG	YNLYARG	ILKQEH	LRPLPG	APLAENH	FPETVLT	LGDSH	
		360	370	380	390	400	410				

The complete length ORF128a nucleotide sequence <SEQ ID 831> is:

1	ATGCAAGCTG	TCCGATACAG	ACCGGAAATT	GACGGATTGC	GGGCCGTGCG
51	CGTGCTATCC	GTCATGATTT	TCCACCTGAA	TAACCGCTGG	CTGCCCGGAG
101	GATTCCTGGG	GGTGGACATT	TTCTTTGTCA	TCTCAGGATT	CCTCATTACC
151	GGCATCATTC	TTTCTGAAAT	ACAGAACGGT	TCTTTTCTT	TCCGGGATTT
201	TTATACCCGC	AGGATTAAGC	GGATTTATCC	TGCTTTTATT	GCGGCCGTGT
251	CGCTGGCTTC	GGTGATTGCC	TCTCAAATCT	TCCTTTACGA	AGATTTCAAC
301	CAATGCGGA	AAACCGTGGA	GCTTTCTGCG	GTTTCTTGT	CCAATATTTA
351	TCTGGGGTTT	CAGCAGGGGT	ATTTGCGATT	GAGTGCCGAC	GAGAACCCCG
401	TACTGCATAT	CTGGTCTTTG	GCAGTAGAGG	AACAGTATTA	CCTCCTGTAT
451	CCTCTTTTGC	TGATATTTTG	CTGCAAAAAA	ACAAAATCGC	TACGGGTGCT
501	GCGTAACATC	AGCATCATCC	TATTTCTGAT	TTTGACTGCC	ACATCGTTTT
551	TGCCAAGCGG	GTTTTATACC	GATATTCTCA	ACCAACCCAA	TACTTATTAC
601	CTTTGACAC	TGAGGTTTCC	CGAGCTGTTG	GCAGGTTTCG	TGCTGGCGGT
651	TTACGGGCAA	ACGCAAAACG	GCAGACGGCA	AACAGCAAAT	GGAAAACGGC
701	AGTTGCTTTC	ATCACTCTGC	TTCGGCGCAT	TGCTTGCTCG	CCTGTTCCGT
751	ATTGACAAAC	ACAATCCGTT	TATCCCGGGA	ATGACCCTGC	TCCTTCCCTG
801	CCTGCTGACG	GCACTGCTTA	TCCGGAGTAT	GCAATACGGG	ACACTTCCGA
851	CCCGCATCCT	GTCGGCAAGC	CCCATCGTAT	TTGTCGGCAA	AATCTCTTAT
901	TCCCTATACC	TGTACCATTG	GATTTTATT	GCTTTCGCCC	ATTACATTAC
951	AGGCGACAAA	CAGCTCGGAC	TGCCCTGCCG	ATCGGCGGTT	GCCGCGTTGA
1001	CGGCCGGATT	TTCCCTGTTG	AGTTATTATT	TGATTGAACA	GCCGCTTAGA
1051	AAACGGAAGA	TGACCTTCAA	AAAGGCATTT	TTCTGCCTCT	ATCTCGCCCC
1101	GTCCCTGATA	CTTGTCGGTT	ACAACCTGTA	CGCAAGGGGG	ATATGAAAC
1151	AGGAACACCT	CCGCCCGTTG	CCCGGCGCGC	CCCTTGCTGC	GGAAAATCAT
1201	TTTCCGGAAA	CCGTCTGAC	CCTCGGCGAC	TCGCACGCGG	GACACCTGCG
1251	GGGGTTTCTG	GATTATGTCG	GCAGCCGGGA	AGGGTGGAAG	GCCAAAATCC
1301	TGTCCCTCGA	TTCCGAGTGT	TTGGTTTGGG	TAGATGAGAA	GCTGGCAGAC
1351	AACCCGTTAT	GTCGAAAATA	CCGGGATGAA	GTTGAAAAG	CCGAAGCCGT
1401	TTTCATTGCC	CAATCTATG	ATTTGAGGAT	GGGCGGCCAG	CCCGTGCCGA
1451	GATTTGAAGC	GCAATCCTTC	CTAATACCGG	GTTTCCAGC	CCGATTGAGG
1501	GAAACCGTCA	AAAGGATAGC	CGCCGTCAAA	CCCGTCTATG	TTTTTGCAAA
1551	CAACACATCA	ATCAGCCGTT	CGCCCCTGAG	GGAGGAAAAA	TTGAAAAGAT
1601	TTGCCGCAAA	CCAATATCTC	CGCCCCATTC	AGGCTATGGG	CGACATCGGC
1651	AAGAGCAATC	AGGCGGTCTT	TGATTTGATT	AAAGATATTG	CCAATGTGCA
1701	TTGGGTGGAC	GCACAAAAAT	ACCTGCCCAA	AAACACGGTC	GAAATATACG
1751	GCCGCTATCT	TTACGGCGAC	CAAGACCACC	TGACCTATTT	CGGTTCTTAT
1801	TATATGGGGC	GGGAATTTCA	CAACACGAA	CGCCTGCTTA	AATCTTCTCG
1851	CGACGGCGCA	TTGCAGTAG			

This encodes a protein having amino acid sequence <SEQ ID 832>:

1	MQAVRYRPEI	DGLRAVAVLS	VMIFHLNNRW	LPGGFLGVDI	FFVISGFLIT
51	GIILSEIQNG	SFSFRDFYTR	RIKRIYPAFI	AAVSLASVIA	SQIFLYEDFN
101	QMRKTVELSA	VFLSNIIYLG	F	QQGYFDLSAD	ENPVLHIWSL
151	PILLIFCCKK	TKSLRVLRLNI	SIILFLILTA	TSFLPSGFYT	DILNQNTYY
201	LSTLRFPELL	AGSLLAVYQ	TQNGRRQTAN	GKRLSSLC	FGALLACLFV
251	IDKHNPFI	PGMTLLLPCLL	TALLIRSMQY	GTLPTILSAS	PIVFVGKISY
301	SLYLYHWIFI	AFAYHITGDK	QLGLPAVSAV	AALTAGFSL	SYLLIEQPLR
351	KRKMTFKKAF	FCLYLAPSLI	LVGYNLYARG	ILKQEHRLPL	PGAPLAENH
401	FPETVLTLD	SHAGHLRGFL	DYVGSREGWK	AKILSLDSEC	LVWVDEKLAD
451	NPLCRKYRDE	VEKAEAVFIA	QFYDLRMGGQ	PVPRFEAQSF	LIPGFPARFR

501 ETVKRIAARK PVYVFANNTS ISRSPLREEK LKRFANQYL RPIQAMGDIG  
 551 KSNQAVFDLI KDIPNVHWVD AQKYLKNTV EIYGRYLYGD QDHLTYFGSY  
 601 YMGREFHKHE RLLKSSRDGA LQ\*

ORF128a and ORF128-1 show 99.5% identity in 622 aa overlap:

5	orf128a.pep	MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG
	orf128-1	MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG
10	orf128a.pep	SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG
	orf128-1	SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG
	orf128a.pep	QQGYFDLSADENPVLHIWISLAVEEQYLLYPLLLIFCCKKTKSLRVLNISIILFLILTA
15	orf128-1	QQGYFDLSADENPVLHIWISLAVEEQYLLYPLLLIFCCKKTKSLRVLNISIILFLILTA
	orf128a.pep	TSFLPSGFYTDILNQPNYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGKRQLSSLC
	orf128-1	SSFLPSGFYTDILNQPNYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGKRQLSSLC
20	orf128a.pep	FGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPIVFGKISY
	orf128-1	FGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPIVFGKISY
25	orf128a.pep	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYLIEQPLRKRKMTFKKAF
	orf128-1	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYLIEQPLRKRKMTFKKAF
30	orf128a.pep	FCLYLAPSLILVGYNLYARGILKQEHRLPLPGAPLAAENHFPETVLTGLGDSHAGHLRGFL
	orf128-1	FCLYLAPSLILVGYNLYARGILKQEHRLPLPGAPLAAENHFPETVLTGLGDSHAGHLRGFL
	orf128a.pep	DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
35	orf128-1	DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
	orf128a.pep	PVPRFEAQSFILPGFPARFRETVKRIAARKPVYVFANNTSISRSPLREEKLRKRFANQYL
40	orf128-1	PVPRFEAQSFILPGFPARFRETVKRIAARKPVYVFANNTSISRSPLREEKLRKRFANQYL
	orf128a.pep	RPIQAMGDIGKSNQAVFDLIKDIPNVHWVDAQKYLKNTVEIYGRYLYGDQDHLTYFGSY
	orf128-1	RPIQAMGDIGKSNQAVFDLIKDIPNVHWVDAQKYLKNTVEIYGRYLYGDQDHLTYFGSY
45	orf128a.pep	YMGREFHKHERLLKSSRDGALQX
	orf128-1	YMGREFHKHERLLKSSHGALQX

#### Homology with a predicted ORF from *N.gonorrhoeae*

50 ORF128 shows 93.4% identity over 244 aa overlap with a predicted ORF (ORF128ng) from *N. gonorrhoeae*:

	orf128.pep	VSLASVIASQIFLYEDFNQMRKTVELSAVF	30
	orf128ng	ILSEIQNGSFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTIELSTVF	112
55	orf128.pep	LSNIYLGFGQGYFDLSADENPVLHIWISLAVEEQYLLYPLLLIFCCKKTKSLRVLNISI	90
	orf128ng	LSNIYLGFRGLGYFDLSADENPVLHIWISLAVEEQYLLYPLLLIFCYKTKSLRVLNISI	172
60	orf128.pep	ILFLILTASSFLPSGFYTDILNQPNYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGK	150
	orf128ng	ILFLILTASSFLPAGFYTDILNQPNYYLSTLRFPPELLVGSLLAVYGQTQNGRRQTENGK	232
	orf128.pep	RQLSSSLCFGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI	210
65	orf128ng	RQLSSLLCFGALLVCLFVIDKHDPFIPGITLLLPCLLTALLIRSMQYGTLPTRILSASPI	292

orf128.pep VFGKISYSLYLYHWIFIAFAPLIRGGKQLGLPA 244  
 |||||  
 orf128ng VFGKISYSLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKR 352

5 The complete length ORF128ng nucleotide sequence <SEQ ID 833> is:

1 ATGCAAGCTG TCCGATACAG GCCTGAAATT GACGGATTGC GGGCCGTCGC  
 51 CGTGTATACC GTCATTATTT TCCACCTGAA TAACCGCTGG CTGCCCCGAG  
 101 GATTCTCTGG GGTGGACATT TTCTTTGTCA TCTCGGGATT CCTCATTACC  
 151 AACATCATT TTTCTGAAAT ACAGAACGGT TCTTTTCTT TCCGGGATTT  
 10 201 TTATACCCGC AGGATTAAGC GGATTATCC TGCTTTTATT GCGCCGTGT  
 251 CCCTGGCTTC GGTGATTGCT TCTCAAATCT TCCTTTACGA AGATTTCAAC  
 301 CAAATGAGGA AAACCATAGA GCTTTCTACG GTTTTTTTGT CCAATATTTA  
 351 TTTGGGGTTC CGATTGGGGT ATTTGATTGTT GAGTGCCGAC GAGAACCCCG  
 401 TACTGCATAT CTGGTCTTTG GCGGTAGAGG AACAGTATTA CCTCCTGTAT  
 15 451 CCTCTTTTGC TGATATTCTG TTACAAAAA ACCAAATCAC TACGGGTGCT  
 501 GCGTAATATC AGCATCATCC TGTTTCTGAT TTTGACCGCA TCATCGTTTT  
 551 TGCCGGCCCG GTTTTATACC GACATCCTCA ACCAACCCaa TACTTATTAC  
 601 CTTTCGACAC TGAGGTTTCC CGAGCTGTTG GTGGGTTTCG TGTTGGCGGT  
 651 TTACGGGCAA ACGCAAACG GCAGACGGCA AACAGAAAT GGAACCGGC  
 20 701 AGTTGCTTTC ATTACTCTGT TTCGGCGCat tgCTTGTCTG CCTGTTCGTG  
 751 ATCGACAAAC ACGATCCGTT TATCCCGGGA ATAACCTTGC TCCTTCCCTG  
 801 CCTGCTGACG GCGCTGCTTA TCCGGAGTAT GCAATACGGG ACACTCCGA  
 851 CCCGCATCCT GTCGGCAAGC CCCATCGTAT TTGTCGGCAA AATCTCTTAT  
 901 TCCCTATACC TGTACCATTG GATTTTTTAT GCCTTCGCCC ATTACATTAC  
 25 951 AGGCGACAAA CAGCTCGGAC TGCTTCCCGT ATCGGCGGTT GCCGCGTTGA  
 1001 CGGCCGGATT TTCCCTGTTG AGCTATTATT TGATTGAACA GCCGCTTGA  
 1051 AAACGGAAGA TGACCTTCAA AAAGGCATT TTCTGCCTTT ATCTCGCCCC  
 1101 GTCCCTGATG CTTGTGCGTT ACAACCTGTA TTCAAGAGGG ATATTGAAAC  
 1151 AGGAACACCT CCGCCCCCTG CCCGGCACGC CCGTTGCTGC GGAAATAAT  
 30 1201 TTTCCGAAA CCGTCTTGAC CCTCGGCGAC TCGCACGCCG GACACCTGCG  
 1251 GGGGTTTCTG GATTATGTCG GCGGCAGGGA AGGGTGGA GCTAAAATCC  
 1301 TGTCCCTCGA TTCGGAGTGT TTGGTTTGGG TGGATGAGAA GCTGGCAGAC  
 1351 AACCCTTGT GCCGAAAATA CCGGGATGAA GTTGAAAAAG CCGAAGCTGT  
 1401 TTTCAATTGCC CAATTCTATG ATTTGAGGAT GGGCGGCCAG CCCGTGCCGA  
 35 1451 GATTGAAGC GCAATCTTTC CTGATACCCG GGTTCAAAGC CCGATTGAGG  
 1501 GAAACCGTCA AGAGGATAGC CGCCGTCAA CCTGTATATG TTTTGCAAA  
 1551 CAATACATCA ATCAGCCGTT CTCCCTTGAG GGAGGAAAAA TTGAAAAGAT  
 1601 TTGCTATAAA CCAATACCTC CGGCCTATTC GGGCTATGGG CGACATCGGC  
 1651 AAGAGCAATC AGGCGGTCTT TGATTGGTT AAAGATATTC CCAATGTGCA  
 40 1701 TTGGGTGGAC GCACAAAAT ACCTGCCCAA AAACACGGTC GAAATACACG  
 1751 GACGCTATCT TTACGGCGAC CAAGACCACC TGACCTATT CGGTTCTTAT  
 1801 TATATGGGGC GGAATTTCA CAAACACGAA CGCCTGCTCA AGCATCCCCG  
 1851 AGGCGGCGCA TTGCAGTAG

This encodes a protein having amino acid sequence <SEQ ID 834>:

45 1 MQAVRYRPEI DGLRAVAVLS VIIFHLNNRW LPGAFLGVDI FFVISGFLIT  
 51 NIILSEIQNG SFSFRDFYTR RIKRIYPAFI AAVSLASVIA SQIFLYEDFN  
 101 QMRKTIELST VFLSNIYLG FRLGYFDLSAD ENPVLHIWSL AVEEQYLLY  
 151 PLLLIFCYKK TKSLRVLRLNI SIILFLILTA SSFLPAGFYT DILNQNTYY  
 201 LSTLRFPELL VGSLLAVYQ TQNGRRQTEN GKRQLLSLLC FGALLVCLFV  
 50 251 IDKHDPFIPG ITLLLPCLLT ALLIRSMQYG TLPTRILSAS PIVFVGKISY  
 301 SLYLYHWIFI AFAHYITGDK QLGLPAVSAV AALTAGFSLL SYYLIEQPLR  
 351 KRKMTFFKAF FCLYLAPSLM LVGYNLYSRG ILKQEHRLPL PGTPVAENN  
 401 FPETVLTLDG SHAGHLRGFL DYVGGREGWK AKILSLDSEC LVWVDEKLAD  
 451 NPLCRKYRDE VEKAEAVFIA QFYDLRMGGQ PVPREFEQSF LIPGFKARER  
 501 ETVKRIAARK PVYVFANNTS ISRSPLREEK LKRFAINQYL RPIRAMGDIG  
 55 551 KSNQAVFDLV KDIPNVHVD AQKYLKNTV EIHGRIYLYGD QDHLTYFGSY  
 601 YMGREFHKHE RLLKHSRGA LQ\*

ORF128ng and ORF128-1 show 95.7% identity in 622 aa overlap:

60 orf128-1.pep MQAVRYRPEIDGLRAVAVLSVIIFHLNNRWLPGAFLGVDIFFVISGFLITGIILSEIQNG  
 |||||  
 orf128ng MQAVRYRPEIDGLRAVAVLSVIIFHLNNRWLPGAFLGVDIFFVISGFLITNIILSEIQNG  
 orf128-1.pep SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG  
 |||||  
 65 orf128ng SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTIELSTVFLSNIYLG

```

5  orf128-1.pep  QQGYFDLSADENPVLHIWVSLAVEEQYYLLYPLLLIFCCKKTKSLRVLNRNISIILFLILTA
   orf128ng      RLGYFDLSADENPVLHIWVSLAVEEQYYLLYPLLLIFCYKTKTKSLRVLNRNISIILFLILTA

10 orf128-1.pep  SSFLPSGFYTDILNQPNNTYYLSTLRFPELLAGSLLAVYGQTQNGRRQTANGKRQLLSLSC
   orf128ng      SSFLPAGFYTDILNQPNNTYYLSTLRFPELLVGSLLAVYGQTQNGRRQTENGKRQLLSLSC

15 orf128-1.pep  FGALLACLFVIDKHNPFPFGMTLLLPCLLTALLIRSMQYGTLPTRILSASPIVEFVGKISY
   orf128ng      FGALLVCLFVIDKHDFPFGITLLLPCLLTALLIRSMQYGTLPTRILSASPIVEFVGKISY

20 orf128-1.pep  SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYLLIEQPLRKRKMTFFKAF
   orf128ng      SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYLLIEQPLRKRKMTFFKAF

25 orf128-1.pep  FCLYLAPSLILVGYNLYARGILKQEHRLPLPGAPLAAENHFPETVLTGLGDSHAGHLRGFL
   orf128ng      FCLYLAPSLMLVGYNLYSRGILKQEHRLPLPGTPVAENHFPETVLTGLGDSHAGHLRGFL

30 orf128-1.pep  DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
   orf128ng      DYVGGREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ

35 orf128-1.pep  PVPRFEAQSFILIPGFPARFRET VKRIA AVKPVYVFANNTSISRSPLREEKLKRFAANQYL
   orf128ng      PVPRFEAQSFILIPGFKARFRET VKRIA AVKPVYVFANNTSISRSPLREEKLKRFAINQYL

40 orf128-1.pep  RPIQAMGDIGKSNQAVFDLIKDI PNVHWVDAQYLPKNTVEIYGRYLYGDQDHLTYFGSY
   orf128ng      RPIRAMGDIGKSNQAVFDLVKDI PNVHWVDAQYLPKNTVEIHGRYLYGDQDHLTYFGSY

45 orf128-1.pep  YMGREFHKHERLLKSSHGGALQX
   orf128ng      YMGREFHKHERLLKHSRGGALQX
                        610      620

```

In addition, ORF218ng shows homology to a hypothetical *H. influenzae* protein:

```

40 sp|P43993|Y392_HAEIN  HYPOTHETICAL PROTEIN HI0392 >gi|1074385|pir||B64007
   hypothetical protein HI0392 - Haemophilus influenzae (strain Rd KW20)
   >gi|1573364 (U32723) H. influenzae predicted coding region HI0392 [Haemophilus
   influenzae] Length = 245
   Score = 239 bits (604), Expect = 3e-62
   Identities = 124/225 (55%), Positives = 152/225 (67%), Gaps = 1/225 (0%)

45 Query: 38  VDIFVISGFLITNIILSEIQNGSFSFRDFYTRRIKRIYPXXXXXXXXXXXXXXXXXXFLYE 97
   +DIFVISGFLIT II++EIQ SFS + FYTRRIKRIYP F+Y
   Sbjct: 1  MDIFVISGFLITGIIITEIQNSFSLSKQFYTRRIKRIYPAFITVMALVSFIASAIIFIYN 60

50 Query: 98  DFNQMRKTIELSTVFLSNYILGFERLGYFDLSADENPVLHIWVSLAVEEQXXXXXXXXXIFC 157
   DFN++RKTIEL+ FLSN YLG GYFDLSA+ENPVLHIWVSLAVE Q I
   Sbjct: 61  DFNKLRTIELAIAFLSNFYLGTLTGQYFDLSANENPVLHIWVSLAVEGQYYLIFPLILILA 120

55 Query: 158 YKKTSLRVLNRNISIILFLILTASSFLPAGFYTDILNQPNNTYYLSTLRFPELLVGSLLAV 217
   YKK + ++VL I++ILF IL A+SF+ A FY ++L+QPN YYLS LRFPELLVGSLLA+
   Sbjct: 121 YKKFREVKVLFITLILFFILLATSFVSANFYKEVLHQPNIYYLSNLRPELLVGSLLAI 180

60 Query: 218 YGQTQNGRRQTENGKRQLLSLLCFGALLVCLFVIDKHDFPFGIT 262
   Y N + Q +L++L L CLF+++ + FIPGIT
   Sbjct: 181 YHNLSN-KVQLSKQVNNILAILSTLLLFSCFLMNNNIAFIPGIT 224

```

This analysis, including the identification of several putative transmembrane domains, suggests that these proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

**Example 99**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 835>:

```

1  ..ATTATTACG AATACCGCTG GATGTTTCTT TACGGCGCAC TGACGACCTT
51  GGGGCTGACG GTCGTGGCAA C.GCGGGCGG TTCGGTATTG GGTCTGTTGT
5  101  TGGCGTTGGC GCGCCTGATT CACTTGGAAA AAGCCGCTGC GCCGATGCGC
151  GTGCTGGCGT GGGCGTTGCG TAAAGTTTCG CTGCTGTATG TTACGCTGTT
201  CCGGGGTACG CCGCTGTTT TGCAGATTGT GATTGCGCG TATGTGTGGT
251  TTCCGTTTTT CGTC..

```

This corresponds to the amino acid sequence <SEQ ID 836; ORF129>:

```

10  1  ..IIYEYRWMFL YGALTTLGLT VVAXAGGSVL GLLLALARLI HLEKAGAPMR
51  VLAWALRKVS LLYVTLFRGT PLFVQIVIWA YVWFPPFV..

```

Further work revealed the complete nucleotide sequence <SEQ ID 837>:

```

1  ATGGATTTTC GTTTGTACAT TATTTACGAA TACCGCTGGA TGTTCCTTTA
51  CGGCGCACTG ACGACCTTGG GGCTGACGGT CGTGGCAACG GCGGGCGGTT
15  101  CCGTATTGGG TCTGTTGTG GCGTTGGCGC GCCTGATTCA CTGGAAAAA
151  GCCGGTGCGC CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AAGTTTCGCT
201  GCTGTATGTT ACGCTGTTCC GGGGTACGCC GCTGTTGTG CAGATTGTGA
251  TTTGGGCGTA TGTGTGGTTT CCGTTTTTCG TCCATCCTTC AGACGGCATT
301  TTGGTCAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT
20  351  GATTGCCGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG
401  AGATTTTCCG CGCGGCCATC CAGTCTATAG ACAAGGACA GATGGAGGCG
451  GCGCGTTCTT TGGGGCTGAC CTATCCGAG GCGATGCGCT ATGTGATTCT
501  GCCGCAGGCA TTGCGCCGCA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA
551  CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GCGGGAGTTG
25  601  GCGTATGTT AGAATACGAT TACGGGCCGG TATTCGGTTT ATGAAGAACC
651  GCTTTACACC GTCGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT
701  GGATATTCTT GCGTTTGAA AAACGTTACA ATCCGAACA CCGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 838; ORF129-1>:

```

30  1  MDFRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
51  AGAPMRVLAW ALRKVSLLYV TFRGTPLEFV QIVWAYVWF PFFVHPSDGI
101  LVSGEAAIAL RRGYGPLIAG SLALIANSQA YICEIFRAGI QSIDKGQMEA
151  ARSLGLTYPQ AMRYVILPQA LRRMLPPLAS EFITLLKDSS LLSVIAVAEL
201  AYWQNTITGR YSVYEEPLYT VALIYLMTT FLGWIFLRLE KRYNPQHR*

```

Computer analysis of this amino acid sequence gave the following results:

### 35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF129 shows 98.9% identity over a 88aa overlap with an ORF (ORF129a) from strain A of *N.*

*meningitidis*:

```

40  orf129.pep      10      20      30      40      50
      IIYEYRWMFLYGALTTLGLTVVAXAGGSVLGLLLALARLIHLEKAGAPMRVLAW
      |||
orf129a  MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
      10      20      30      40      50      60

45  orf129.pep      60      70      80
      ALRKVSLLYVTLFRGTPLEFVQIVWAYVWFPPFV
      |||
orf129a  ALRKVSLLYVTLFRGTPLEFVQIVWAYVWFPPFVHPSDGI LVSGEAAIALRRYGPLIAG
      70      80      90      100     110     120

50  orf129a  SLALIANSQA YICEIFRAGI QSIDKGQMEA ARSLGLTYPQ AMRYVILPQA LRRMLPPLAS
      130     140     150     160     170     180

```

The complete length ORF129a nucleotide sequence <SEQ ID 839> is:

```

1  ATGGATTTTC GTTTGTACAT TATTTACGAA TACCGCTGGA TGTTCCTTTA
51  CGGCGCACTG ACGACCTTGG GGCTGACGGT CGTGGCGACG GCGGGCGGTT

```



-460-

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70

101 CGGTATTGGG TCTGTTGTTG GCGTTGGCGC GCCTGATTCA CTTGGAAAAA  
151 GCCGGTGCGC CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AGGTTTCGCT  
201 GCTGTATGTT ACGCTGTTC GGGGTACGCC GCTGTTTGTG CAGATTGTGA  
251 TTTGGGCGTA TGTGTGGTTT CCGTTTTTCG TCCATCCTTC AGACGCGATT  
301 TTGTTAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT  
351 GATTGCCGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG  
401 AGATTTTCCG CGCGGGCATC CAGTCTATAG ACAAAGGACA GATGGAGGCG  
451 GCGCGTTCTT TGGGGCTGAC CTATCCGCAG GCGATGCGCT ATGTGATTCT  
501 GCCGCAGGCA TTGCGCCGTA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA  
551 CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GGCGGAGTTG  
601 GCGTATGTT AGAATACGAT TACGGGCCGG TATTCGGTTT ATGAAGAACC  
651 GCTTTACACC GTCGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT  
701 GGATATTCTT GCGTTTGAA AACGTTACA ATCCGCAACA CCGCTGA

This encodes a protein having amino acid sequence <SEQ ID 840>:

15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70

1 MDFRFDIIE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK  
51 AGAPMRVLAW ALRKVSLLYV TLFRTPLFV QIVIWAYVWF PFFVHPSDGI  
101 LVSGEAAIAL RRGYGPIAG SLALIANSQA YICEIFRAGI QSIDKGQMEA  
151 ARSLGLTYPQ AMRYVILPQA LRRMLPLAS EFITLLKDSS LLSVIAVAEL  
201 AYSQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR\*

20 ORF129a and ORF129-1 show 100.0% identity in 248 aa overlap:

25  
30  
35  
40  
45  
50  
55  
60  
65  
70

orf129a.pep MDFRFDIIEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW  
orf129-1 MDFRFDIIEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW  
orf129a.pep ALRKVSLLYVTLFRGTPLFVQIVIWAYVWF PFFVHPSDGI LVSGEAAIALRRGYGPIAG  
orf129-1 ALRKVSLLYVTLFRGTPLFVQIVIWAYVWF PFFVHPSDGI LVSGEAAIALRRGYGPIAG  
orf129a.pep SLALIANSQAYICEIFRAGIQSIDKGQMEAAARSLGLTYPQAMRYVILPQALRRMLPLAS  
orf129-1 SLALIANSQAYICEIFRAGIQSIDKGQMEAAARSLGLTYPQAMRYVILPQALRRMLPLAS  
orf129a.pep EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE  
orf129-1 EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE  
orf129a.pep KRYNPQHRX  
orf129-1 KRYNPQHRX

#### Homology with a predicted ORF from *N.gonorrhoeae*

ORF129 shows 98.9% identity over a 88 aa overlap with a predicted ORF (ORF129ng) from *N.gonorrhoeae*:

45  
50  
55  
60  
65  
70

orf129.pep IIEYRWMFLYGALTTLGLTVVAXAGGSVLGLLLALARLIHLEKAGAPMRVLAW 54  
orf129ng MDFRFDIIEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW 60  
orf129.pep ALRKVSLLYVTLFRGTPLFVQIVIWAYVWF PFFV 88  
orf129ng ALRKVSLLYVTLFRGTPLFVQIVIWAYVWF PFFVILHTAFLGNAMRQSRRVPDKGRWIAG 120

An ORF129ng nucleotide sequence <SEQ ID 841> was predicted to encode a protein having amino acid sequence <SEQ ID 842>:

55  
60  
65  
70

1 MDFRFDIIE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK  
51 AGAPMRVLAW ALRKVSLLYV TLFRTPLFV QIVIWAYVWF PFFVILHTAF  
101 LGNAMRQSRR VPDKGRWIAG SLELNCQPRG RKTRGEFFPG ESNLGTPEPRN  
151 PLMSGQRRFP GCENWYPPQN FIKK\*

Further work revealed the following gonococcal sequence <SEQ ID 843>:

1 ATGGATTTTc gTTTGACAT TATTTAcgaA TACCGCTGGA TGTTTCTTTA

-461-

51 CGGCGCACTG Acgaccttgg ggctgacggt cgtggcgacg gCGGGCGGTT  
 101 CGGtatttgg TCTGTTGTTG GCGTTGGCGC GCCTGATTCA CTTGGAAAAA  
 151 GCGGTTGCGC CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AGGTTTCGCT  
 201 GCTGTACGTT ACCCTGTTCC GGGGTACGCC GCTGTTTGTG CAGATTGTGA  
 5 251 TTTGGGCGTA TGTGTGTTT CCGTTTTTCG TCCATCCTTC AGACGGCATT  
 301 TTGGTCAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT  
 351 GATTGCCGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG  
 401 AGATTTTCCG CGCGGGCATC CAGTCTATAG ACAAAGGACA GATGGAGGCG  
 10 451 GCGTGTCTT TGGGACTGAC CTATCCGCAG GCGATGCGCT ATGTGATTCT  
 501 GCCGAGGCA TTGCGCCGTA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA  
 551 CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GCGGAGATTG  
 601 GCGTATGTT AGAATACGAT TACGGGCCGG TATTCGGTTT ATGAAGAACC  
 651 GCTTTACACC GCCGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT  
 701 GGATATTCCT GCGTTTGAA AAACGTTACA ATCCGCAACA CCGCTGA

15 This corresponds to the amino acid sequence <SEQ ID 844; ORF129ng-1>:

1 MDRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK  
 51 AGAPMRVLAW ALRKVSLLYV TFRGTPLFV QIVIWAYVWF PFFVHPSDGI  
 101 LVSGEAAIAL RRGYGPLIAG SLALIANSQA YICEIFRAGI QSIDKGQMEA  
 151 ARSLGLTYPQ AMRYVILPQA LRRMLPLAS EFITLLKDSS LLSVIAVAEL  
 201 AYVQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR\*

ORF129ng-1 and ORF129-1 show 99.2% identity in 248 aa overlap:

orf129-1.pep MDRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW  
 25 orf129ng-1 MDRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW  
 orf129-1.pep ALRKVSLLYVTLFRGTPLFVQIVIWAYVWF PFFVHPSDGI LVSGEAAIALRRGYGPLIAG  
 orf129ng-1 ALRKVSLLYVTLFRGTPLFVQIVIWAYVWF PFFVHPSDGI LVSGEAAIALRRGYGPLIAG  
 30 orf129-1.pep SLALIANSQAYICEIFRAGIQSIDKGQMEAAARSLGLTYPQAMRYVILPQALRRMLPLAS  
 orf129ng-1 SLALIANSQAYICEIFRAGIQSIDKGQMEAAARSLGLTYPQAMRYVILPQALRRMLPLAS  
 35 orf129-1.pep EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE  
 orf129ng-1 EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTAAALIYLLMTTFLGWIFLRLE  
 orf129-1.pep KRYNPQHRX  
 40 orf129ng-1 KRYNPQHRX

In addition, ORF129ng-1 is homologous to an ABC transporter from *A.fulgidus*:

2650409(AE001090) glutamine ABC transporter, permease protein (glnP)  
 [Archaeoglobus fulgidus] length = 224  
 Score = 132 bits (329), Expect = 2e-30  
 45 Identities = 86/178 (48%), Positives = 103/178 (57%), Gaps = 18/178 (10%)  
 Query: 65 VSLLYVTLFRGTPLFVQIVIWAYVWF PFFVHPSDGI LVSGEAAIALRRGYGPLIAGSLAL 124  
 +S YV + RGTP L VQI+I +F P+ GI + E A G +AL  
 Sbjct: 58 ISTAYVEVIRGTPLLVQILI-----VYFGLPAIGINLQPEPA-----GIIAL 99  
 50 Query: 125 IANSQAYICEIFRAGIQSIDKGQMEAAARSLGLTYPQAMRYVILPQALRRMLPLASEFIT 184  
 SGAYI EI RAGI+SI GOMEAA SLG+TY QAMRYVI PQA R +LP L +EFI  
 Sbjct: 100 SICSGAYIAEIVRAGIESIPIGQMEAAARSLGMTYLOAMRYVIFPQAFRNILPALGNEFIA 159  
 55 Query: 185 LLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTAAALIYLLMTTFLGWIFLRLEKR 242  
 LLKDSSLLSVI++ EL V I P AL YL+MT L + +K+  
 Sbjct: 160 LLKDSSLLSVISIVELTRVGRQIVNTTFNAWTPFLGVALFYLMMTIPLSRLVAYSQKK 217

This analysis, including the identification of transmembrane domains in the two proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful  
 60 antigens for vaccines or diagnostics, or for raising antibodies.

**Example 100**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 845>:

```

1  ..CTGAAAGAAT GCCGTCTGAA AGACCCTGTT TTTATTCCAA ATATCGTTTA
51 TAAGAACATC GCCATTACTT TCCTGCTCTT GCACGCCGCC GCCGAACTTT
5  101 GGCTGCCCGC GCAAACCGCC GGTTTTACCG CGCTCGCCGT CGGCTTCATC
151 CTGCTCGCCA AGCTGCGTGA gCTTCACCAT CACGAACTCT TACGTAAACA
201 cTACGTCCGC ACTTATTACy TGCTCCAAC TTTGCGGCC GCAGgcTAgT
251 TTGTGGACAG GCGCGCGGwA ATTACAAAAC CTGCCCGCyT CCGCGCCCT
10  301 GCACCTGATT ACCCTCGGCG GCATGATGGG CGGCGTGATG ATGGTGTGGc
351 TGACCGCCGG ACTGTGGCAC AGCGGCTTTA CCAAACTCGA CTACCCCAA
401 CTCTGCCGCA TTGCCGTCCC CATCCTTTTC GCCGCCGCCG TCTCGCGCGC
451 TTTCTTGtTG AACGTGAACC CGrTATTTT CATTACCGTT CTTGCGATTC
501 TGACCGCCGC CGTATTCGTA CTGTATCTTT TCrCGTTTAT ACCGATATT
551 CGGGCGAATG CGTTTACAGA CGATCCGGAr TAR

```

15 This corresponds to the amino acid sequence <SEQ ID 846; ORF130>:

```

1  ..LKECRLKDPV FIPNIVYKNI AITFLLLHAA AELWLPAQTA GFTALAVGFI
51  LLAKLRELHH HELLRKHYVR TYLLQLFAA AGSLWTGAAX LQNLPASAPL
101  HLITLGGMMG GVMVWLTAG LWHSGFTKLD YPKLCRIAVP ILFAAAVSR
151  FLXNVNPXFF ITVPAILTAA VFVLYLFXFI PIFRANAFTD DPE*

```

20 Further work revealed the complete nucleotide sequence <SEQ ID 847>:

```

1  ATGCGGCCGT TTTTCGTCGG CGCGCGGGTG CTGCCATAC TCGGTGCGCT
51  GGTGTTTTTC ATCAACCCCG GTGCCATCGT CCTGCACCGC CAAATTTTCT
101  TGGAACTTAT GCTGCCGCGC GCATACGGCG GTTTTTTGAC TCGCGCTTTG
151  TTGGACTGGA CGGGTTTTTC GGGTAACCTG AAACCTGTCG CGACTTTGAT
25  201 GCGCGCATTA TTGCTCGCCG CATCCGCTAT ACTGCCCTTT TCGCCGCAAA
251  CTGCCCTCGTT TTTGCTCGCC GCCTATTGGC TGGTGTGCT GCTGTTCTGC
301  GCCCGGCTGA TTTGGCTAGA CCGAAACACC GACAACTCG CCCTGCTAAT
351  GTTACTTGCC GCGTTCCTG TTTTTCAGAC GGCATATGCC GTCAGCGGCG
401  ATTTGAACCT GTTGC GCGCG CAAGTGCATC TAAATATGGC GGCGGTGATG
30  451 TTCGTATCCG TCGCGCTCAG TATTCTTTG GCGCGGAAG CCCTGAAAGA
501  ATCCGCTCTG AAAGACCTG TTTTATTCC AAATATCGTT TATAAAACA
551  TCGCCATTAC TTTCTGCTC TTGCACGCCG CCGCCGAAC TTGGCTGCCC
601  GCGCAAACCG CCGGTTTTAC CGCGCTCGCC GTCGGCTTCA TCCTGCTCGC
651  CAAGCTGCGT GAGCTTCACC ATCAGCAACT CTTACGTAA CACTACGTCC
35  701 GCACCTATTA CTTGCTCCAA CTCTTTGCCG CCGCAGGCTA TTTGTGGACA
751  GGCGCGCGCA AATTACAAA CCTGCCGCC TCCGCGCCCC TGACCTGAT
801  TACCCTCGGC GGCATGATGG GCGGCGTGAT GATGGTGTGG CTGACCGCCG
851  GACTGTGGCA CAGCGGCTTT ACCAACTCG ACTACCCCAA ACTCTGCGCG
901  ATTGCCGTCC CCATCCTTTT CGCCGCCGCC GTCTCGCGCG CTTTCTTGAT
40  951 GAACGTGAAC CCGATATTTT TCATTACCGT TCCTGCGATT CTGACCGCGC
1001 CCGTATTCGT ACTGTATCTT TTCACGTTTA TACCGATATT TCGGGCGAAT
1051 GCGTTTACAG ACGATCCGGA ATAA

```

This corresponds to the amino acid sequence <SEQ ID 848; ORF130-1>:

```

1  MRPFVGA AV LAILGALVFF INPGAIVLHR QIFLEMLPA AYGGFLTAAL
45  51 LDWTGFSGNL KPVATLMAAL LLAASAILPF SPQTASEFFVA AYWLVLLEFC
101  ARLIWLD RNT DNFALLMLLA AFTVFQTAYA VSGDLNLLRA QVHLNMAAVM
151  FVSVRSILL GAEALKECRL KDPVFIPNIV YKNIAITFLL LHAAAEWLWP
201  AQTAGFTALA VGFILLAKLR ELHHHELLRK HYVRTYLLQ LFAAAGYLWT
251  GAAKLQNLPA SAPLHLITLG GMMGGVMMVW LTAGLWHS GF TKLDYPKLCR
50  301 IAVPILFAAA VSRAFLMNVN PIFFITVPAI LTA AVFVLYL FTFIPIFRAN
351  AFTDDPE*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF130 shows 94.3% identity over a 193aa overlap with an ORF (ORF130a) from strain A of *N.*

55 *meningitidis*:

			10	20	30
	orf130.pep	LKECRLKDPVFIPNIVYKNIAITFLLHAA      :			
5	orf130a	LNL <sup>140</sup> LR <sup>150</sup> AQVHLNMAAVMFVSVRVSI <sup>160</sup> LGAEALKECRLKDPVFIPNVVYKNIAITFLLHAA <sup>170</sup> <sup>180</sup> <sup>190</sup>			
		40	50	60	70
	orf130.pep	AELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLQLFAAAGSLWTGAAX      :	80	90	
10	orf130a	AELWLPAQTAGFTSLAVGFILLAKLRELHHHELLRKHYVRTYYLLQLFAAAGYLWTTGAAK <sup>200</sup> <sup>210</sup> <sup>220</sup> <sup>230</sup> <sup>240</sup> <sup>250</sup>			
		100	110	120	130
	orf130.pep	LQNLPASAPLHLITLGGMMGVVMVWLTAGLWHSGFTKLDYPKLCR <sup>140</sup> IAVPILFAAAVSRA      :	140	150	
15	orf130a	LQNLPASAPLHLITLGGMGSVMVWLTAGLWHSGFTKLDYPKLCR <sup>160</sup> IAVPILFAAAVSRA <sup>260</sup> <sup>270</sup> <sup>280</sup> <sup>290</sup> <sup>300</sup> <sup>310</sup>			
		160	170	180	190
	orf130.pep	FLXNVNPXF <sup>170</sup> FITVPAILTAAVFVLYLEXFIPIFRANAFTDDPEX    : :	190		
20	orf130a	VLMNVNPIFFITVPAILTAAVFVLYLLTFVPIFRANAFTDDPEX <sup>320</sup> <sup>330</sup> <sup>340</sup> <sup>350</sup>			

25	1	ATGCGGCCGT	TTTTCGTCGG	CGCGGCGGTG	CTTGCCATAC	TCGGTGCGCT
	51	GGTGTTTTTC	ATCAACCCCG	GTGCCATCGT	CCTGCACCGC	CAAATTTTCT
	101	TGGAACCTAT	GCTGCCGCGC	GCATACGGCG	GTTTTTTGAC	TGCGGCTTTG
	151	TTGGAGCTGA	CGGGTTTTTC	GGGTAACCTG	AAACCTGTGC	CGACTTTGAT
30	201	GGCGGCATTA	TGTCTCGCGC	CATCCGCTAT	ACTGCCCTTT	TCGCGCGAAA
	251	CTGCCCTCGTT	TTTCGTGCGC	GCCTATTGGC	TGGTGTTGCT	GCTGTTCTGC
	301	GCCCGGCTGA	TTTGGCTAGA	CCGAAACACC	GACAAC'TTCG	CCCTGCTAAT
	351	GTTACTTGCC	GCGTTCAC TG	TTTTTCAGAC	GGCATATGCG	GTACGCGGCG
35	401	ATTTGAACCT	TGTGCGCGCG	CAAGTGCATC	TGAATATGCG	GGCGGCTGATG
	451	TTCGTATCCG	TGCGCGTCAG	TATTCTTTTG	GGCGCGGAAG	CCCTGAAAGA
	501	ATGCCG'TCTG	AAAGACCCAG	TATTCATCCC	CAATGTGCTC	TATAAAAAACA
	551	TCGCCATTAC	CTTCCTGCTC	CTGCA'CGCCG	CCGCGCAACT	TTGGCTGCCT
40	601	GCGCAAACCG	CGGTTTTTAC	CTCGCTCGCG	GTCGGCTTTA	TCTCGT'TTGC
	651	CAAGCTGCGT	GAGCTTCACC	ATCAGCAACT	CCTGCGCAAA	CAC'TACGTCC
	701	GCACTTATTA	CCTGCTCCAA	CTCTTTGCCG	CCG'CAGGCTA	TTTGTGGACA
	751	GGCGCGGCGA	AATTACAAAA	CCTGCCCCGC	TCCGCGCCCC	TGCACCTGAT
45	801	TACCCTCGGT	GGCATGATGG	GCAGCGTGAT	GATGGTGTGG	CTGACTGCCG
	851	GACTGTGGCA	CAGCGGCTTT	ACCAAGCTCG	ACTACCCGAA	ACTCTGCCGC
	901	ATCGCCGTCC	CCATCCTNTT	CGCCGCCGCC	GTTTCGCGCG	CTGTTTTAAT
	951	GAACGTA'AAAC	CCGATATTCT	TCATCACCGT	CCCCGCAATT	CTGACCGCCG
45	1001	CCGTGTTCGT	GCTTTACCTG	CTGACATT'CG	TACCGATCTT	TCGGGCGAAC
	1051	CGGTTTTACG	ACGATCCGGA	ATAA		

[illegible]

	orf130a.pep	MRPFFVGAAVLAILGALVEFFINPGAIVLHRQIFLEIMLPAAYGGFLTAALLDWTGFSGNL
60	orf130-1	MRPFFVGAAVLAILGALVEFFINPGAIVLHRQIFLEIMLPAAYGGFLTAALLDWTGFSGNL
	orf130a.pep	KPVATLMAALLLAASAILPFPQTASFFVAAYWLVLLLFCARLIWLDNRNTDNFALLMLLA
	orf130-1	KPVATLMAALLLAASAILPFPQTASFFVAAYWLVLLLFCARLIWLDNRNTDNFALLMLLA
65	orf130a.pep	AFTVFQYAYAVSGDLNLLRAQVHLNMAAVMFVSVRVSIILGAEALKECRLKDPVFIPNVV

-464-

```

      orf130-1      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
5      orf130a.pep  YKNIAITFLLHAAELWLPAQTAGFTSLAVGFILLAKLRELHHHELLRKHYVRTYLLQ
      orf130-1      YKNIAITFLLHAAELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYLLQ
      orf130a.pep  LFAAAGYLWTGAACKLQNLPAAPLHLITLGGMMGSVMVWLTAGLWHSGETKLDYPKLCR
10     orf130-1      LFAAAGYLWTGAACKLQNLPAAPLHLITLGGMMGSVMVWLTAGLWHSGETKLDYPKLCR
      orf130a.pep  IAVPILFAAAVSRAVLMNVNPIFFITVPAILTAAVFVLYLLTFVPIFRANAFTDDPE
      orf130-1      IAVPILFAAAVSRAFLMNVNPIFFITVPAILTAAVFVLYLLTFPIFRANAFTDDPE
15

```

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF130 shows 91.7% identity over a 193 aa overlap with a predicted ORF (ORF130ng) from *N.gonorrhoeae*:

```

20     orf130.pep      LKECRLKDPVFIPNIVYKNIAITFLLHAA      30
      orf130ng      LNLRAQVHLNMAAVMFVSVRVSVLLGTETLKECRLKDPVFIPNVIYKNIAIT-LLLHAA      201
      orf130.pep  AELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYLLQLFAAAGSLWTGAAX      90
25     orf130ng      AELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYLLQLFAAAGYLWTGAAX      261
      orf130.pep  LQNLPAAPLHLITLGGMMGSVMVWLTAGLWHSGETKLDYPKLCRIAVPILFAAAVSRA      150
      orf130ng      LQNLPAAPLHLITLGGMTGGVMVWLTAGLWHSGETKLDYPKLCRIAVSILFASAVSRA      321
30     orf130.pep  FLXNVNPFITVPAILTAAVFVLYLFXFIPIFRANAFTDDPE      193
      orf130ng      VLMNVNPIFFITVPEILTAAVFMLYLLTFPIFRANAFTDDPE      364

```

An ORF130ng nucleotide sequence <SEQ ID 851> was predicted to encode a protein having amino acid sequence <SEQ ID 852>:

```

40     1  MNKEFFTHPMR PFFVGA AVLA ILGALVFFHQ PRRYHPAPPN FLGTYAAGCI
      51  RRFEDYRFVG PDGFFRQ PET CRYFDGGVVA CCGCFIAVFT ATCRIFRRRL
      101 LAGVA AVLRL ADLARROHRT LRSVDVTA AF TVFQTAYAVS GDLNLLRAQV
      151 HLNMAAVMFV SVRVSVLLGT ETLKECRLKD PVFIPNVIYK NIAITLLLHA
      201 AAELWLPAQT AGFTALAVGF ILLAKLRELH HHELLRKHYV RTYLLQLFA
      251 AAGYLWTGAA KLQNLPAAP LHLITLGGMT GGVMVWLT GLWHSGETKL
      301 DYPKLCRIAV SILFASAVSR AVLNVNPIF FITVPEILTA AVFMLYLLTF
      351 VPIFRANAFT DDPE*

```

Further work revealed the following gonococcal DNA sequence <SEQ ID 853>:

```

45     1  ATGCGCCCGT TTTTCGTCGG TCGCGCAGTA CTTGCCATAC TCGGTGCGTT
      51  GGTGTTTTTT ATCAACCCCG GCGCTATCAT CCTGCACCGC CAAATTTTCT
      101 TGGAATTAT GCTGCCGGCT GCATACGGCG GTTTTTTGAC TACCGCTTTG
      151 TTGGACCGGA CGGGTTTTTC AGGCAACCTG AAACCTGCCG CTACTTTGAT
      201 GCGGTGTTG TTGCTTGTG CCGCTGTTT ATTGCCGTTT TTACCGCAAC
50     251 TTGCCGATT TTTCGTCGCC GCCTATTGGC TGGTGTGCT GCTGTTCTGC
      301 GCCTGGCTGA TTTGGCTCGA CCGCAACACC GACAACTTCG CTCTGTTGAT
      351 GTTACTTGCC GCATTTACCG TTTTTCAGAC GGCCTATGCC GTCAGCGGCG
      401 ATTTGAACTT ACTGCGCGCG CAAGTGCATT TGAATATGGC GGCGGTCATG
      451 TTCTGTATCCG TCCGCTCAG CGTCTTTTGG GGCACGGAAA CCCTGAAAGA
55     501 ATGCCGCTCTG AAAGACCCCG TATTCATCCC CAACGTTATC TATAAAACA
      551 TCGCCATCAC CTGCTGCTG CACGCCGCCG CCGAAGTTTG GCTGCCCGCG
      601 CAAACCGCCG GTTTTACTGC GCTTGCCGTC GGCTTCATCC TGCTCGCCAA
      651 CTTGCGCGAA CTGCACCATC ACGAACTCTT ACGCAAACAC TACGTCCGCA
      701 CTTATTACCT GCTCCAGCTC TTTGCGCGCG CAGGTTATCT GTGGACAGGC
60     751 GCGGCGAAAC TGCAAAACCT GCCCGCCTCC GCGCCCCTGC ACCTGATTAC
      801 CCTCGCGCGC ATGACGGGTG GCGTGATGAT GGTGTGGCTG ACTGCCGGAC
      851 TGTGGCACAG CGGCTTTACC AAACCTCGACT ACCCGAAACT CTGCCGCATC

```

-465-

901 GCGCTCTCCA TCCTTTTCGC CTCGCCCGTT TCGCGCGCTG TTTTAATGAA  
 951 CGTGAATCCG ATATTCTTCA TCACCGTTCC CGAGATTCTG ACCGCCGCCG  
 1001 TGTTTCATGCT TTACCTGCTG ACGTTCGTAC CGATTTTTCG AGCGAACGCC  
 1051 TTACAGACG ATCCGGAATA A

5 This corresponds to the amino acid sequence <SEQ ID 854; ORF130ng-1>:

1 MRPFVGA AV LAILGALVFF INPGA IILHR QIFLELMLPA AYGGFLTTAL  
 51 LDRTGFSGNL KPAATLMAVL LLVAAVLLPF LPQLAAFFVA AYWLVL LLLFC  
 101 AWLIWLD RNT DN FALLMLLA AFTVFQ TAYA VSGDLNLLRA QVHLNMAAVM  
 151 FVSVRVSVLL GTETLKECRL KDPVFIPNVI YKNIAITLLL HAAELWLPA  
 10 TAGFTALAV GFILLAKLRE LHHHELLRKH YVRTYLLQL FAAAGYLWTG  
 251 AAKLQNL PAS APLHLITLGG MTGGVMVWL TAGLWHSGET KLDYPKLCRI  
 301 AVSILFASAV SRAVLMNVNP IFFITVPEIL TAAVFMLYLL TFVPIFRANA  
 351 FTDDPE\*

ORF130ng-1 and ORF130-1 show 92.4% identity in 357 aa overlap:

15	orf130-1.pep	MRPFVGA AV LAILGALVFF INPGA IILHR QIFLELMLPA AYGGFLTTA ALLDWTGFSGNL
	orf130ng-1	MRPFVGA AV LAILGALVFF INPGA IILHR QIFLELMLPA AYGGFLTTA ALLDWTGFSGNL
20	orf130-1.pep	KPVATLMA ALLLAASAILPFSPQTASFFVAAYWLVL LLLFCARLIWLD RNTDN FALLMLLA
	orf130ng-1	KPAATLMA VLLVA AVLLPF LPQLAAFFVAAYWLVL LLLFCAWLIWLD RNTDN FALLMLLA
25	orf130-1.pep	AFTVFQ TAYAVSGDLNLLRA QVHLNMAAVMFVSVRV SILLGAEALKECRL KDPVFIPNIV
	orf130ng-1	AFTVFQ TAYAVSGDLNLLRA QVHLNMAAVMFVSVRV SILLGAEALKECRL KDPVFIPNIV
30	orf130-1.pep	YKNIAITF LLLHAAELWLPA QTAGFTALAVGFILLAKLRELH HHELLRKH YVRTYLLQ
	orf130ng-1	YKNIAIT- LLLHAAELWLPA QTAGFTALAVGFILLAKLRELH HHELLRKH YVRTYLLQ
35	orf130-1.pep	LFAAGYLWTGA AKLQNL PASAPLHLITLGGMTGGVMVWL TAGLWHSGETKLDYPKLCR
	orf130ng-1	LFAAGYLWTGA AKLQNL PASAPLHLITLGGMTGGVMVWL TAGLWHSGETKLDYPKLCR
40	orf130-1.pep	IAVPILFAA AVSRAVLMNVNP IFFITVPA ILTA AVFVLYLFTFIPIFRANA FTDDPEX
	orf130ng-1	IAVSILFASAV SRAVLMNVNP IFFITVPEILTA AVFVLYLFTFIPIFRANA FTDDPEX

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

#### 40 Example 101

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 855>:

1 ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCAAT  
 51 TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA  
 45 101 CCGGCTGGTG TAAGCCGAGA AAACCGGCTG CCATCGATTT TTGGGATATT  
 151 GCGCGCGAGA GTCCGCCGTC TTAGGGGAC TACGAGATAC CGCTTTCAGA  
 201 CCGCAATAGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT  
 251 ACTTTTACAG GAAATAGGG AAGTTTGAAG C.TGCGGGCT GGATTGGCGT  
 301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG GAGGATTGGA  
 351 CTGCTTGGA AAG..

50 This corresponds to the amino acid sequence <SEQ ID 856; ORF131>:

1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAAIDFWDI  
 51 GGESPPSLGD YEIPLSDGNS SVRANEYESA QSYFYRKIG KFEXCGLDWR  
 101 TRDGKPLIET FKQGGFDCLE K..

Further work revealed the complete nucleotide sequence <SEQ ID 857>:

55 1 ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCAAT  
 51 TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA

-466-

5  
101 CCGGCTGGTG TAAGCCGAGA AAACCGGCTG CCATCGATTT TTGGGATATT  
151 GCGGCGGAGA GTCCGCCGTC TTTAGGGGAC TACGAGATAC CGCTTTCAGA  
201 CCGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT  
251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG CCTGCGGGCT GGATTGGCGT  
301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG GAGGATTGTA  
351 CTGCTTGGAA AAGCAGGGGT TCGGGCGCAA CGGTCTGTCC GAGCGCGTCC  
401 GATGGTAA

This corresponds to the amino acid sequence <SEQ ID 858; ORF131-1>:

10  
1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAAIDFWDI  
51 GGESPPSLGD YEIPLSDGNR SVRANEYESA QQSYFYRKIG KFEACGLDWR  
101 TRDGKPLIET FKQGGFDCL E KQGLRRNGLS ERVRW\*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF131 shows 95.0% identity over a 121aa overlap with an ORF (ORF131a) from strain A of *N.*

15 *meningitidis*:

		10	20	30	40	50	60
orf131.pep	MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSSLTGWCKPRKPAAIDFWDIGGESPPSLGD						
orf131a	MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSSLSGWCKPRKPAAIDFWDIGGESPPSLED						
		10	20	30	40	50	60
orf131.pep	YEIPLSDGNSSVRANEYESAQQSYFYRKIGKFEXCGLDWRTRDGKPLIETFKQGGFDCL E						
orf131a	YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQEGFDCLK						
		70	80	90	100	110	120
orf131.pep							
orf131a							
		70	80	90	100	110	120
orf131.pep	K						
orf131a	KQGLRRNGLSERVRW						
	130						

The complete length ORF131a nucleotide sequence <SEQ ID 859> is:

35  
1 ATGGAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCATT  
51 TACGGTTGCA GGCTGCCGGT TGGCAGGTTG GTATGAGTGT TCGTCCCTGT  
101 CCGGCTGGTG TAAGCCGAGA AAACCTGCCG CCATCGATTT TTGGGATATT  
151 GCGGCGGAGA GTCCTCCGTC TTTAGAGGAC TACGAGATAC CGCTTTCAGA  
201 CCGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT  
251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG CCTGCGGGTT GGATTGGCGT  
40  
301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG AAGGTTTGA  
351 TTGTTTGAAG AAGCAGGGGT TCGGGCGCAA CGGTCTGTCC GAGCGCGTCC  
401 GATGGTAA

This encodes a protein having amino acid sequence <SEQ ID 860>:

45  
1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLSGWCKPR KPAAIDFWDI  
51 GGESPPSLED YEIPLSDGNR SVRANEYESA QQSYFYRKIG KFEACGLDWR  
101 TRDGKPLIET FKQEGFDCLK KQGLRRNGLS ERVRW\*

ORF131a and ORF131-1 show 97.0% identity in 135 aa overlap:

orf131a.pep	MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSSLSGWCKPRKPAAIDFWDIGGESPPSLED
orf131-1	MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSSLTGWCKPRKPAAIDFWDIGGESPPSLGD
orf131a.pep	YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQEGFDCLK
orf131-1	YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQGGFDCL E
orf131a.pep	KQGLRRNGLSERVRW

orf131-1 KQGLRRNGLSERVRWX

### Homology with a predicted ORF from *N.gonorrhoeae*

ORF131 shows 89.3% identity over 121 aa overlap with a predicted ORF (ORF131ng) from

5 *N.gonorrhoeae*:

```

orf131.pep  MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD  60
            ||||:|||| |||:||||||| ||:||||||| |||
orf131ng    MEIRVIKYTATAALFAFTVAGCRLAGWYECSSLSGWCKPRKPAAIDFWDIGGESPLSLED  60

10 orf131.pep  YEIPLSDGNSSVRANEYESAQQSYFYRKIGKFEXCGLDWRTRDGKPLIETFKQGGFDCLE  120
            ||||| ||||| |||||:||||| ||||| |||||:| ||| |||||
orf131ng    YEIPLSDGNRSVRANEYESAQKSYFYRKIGKFEACGLDWRTRDGKPLVERFKQEGFDCLE  120

15 orf131.pep  K 121
            |
orf131ng    KQGLRRNGLSERVRW 134

```

A complete length ORF131ng nucleotide sequence <SEQ ID 861> was predicted to encode a protein having amino acid sequence <SEQ ID 862>:

```

20 1 MEIRVIKYTA TAALFAFTVA GCRLAGWYEC LSLSGWCKPR KPAAIDFWDI
    51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
    101 TRDGKPLVER FKQEGFDCLE KQGLRRNGLS ERVRW*

```

Further work revealed the following gonococcal DNA sequence <SEQ ID 863>:

```

25 1 ATGGAAATTC GGGTAATAAA ATATACGGCA ACGGCTGCGT TGTTTGCAAT
    51 TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCTTGT
    101 CCGGCTGGTG TAAGCCGAGA AAACCTGCCG CCATCGATTT TTGGGATATT
    151 GCGGCGGAGA GtccgctGTC TTTAGAGGAC TACGAGATAC CGCTTTCAGA
    201 CGGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCG CAAAATCTT
    251 ACTTTTATAG GAAAATAGGG AAGTTTGAAG CCTGCGGGTT GGATTGGCGT
    301 ACGCGTGACG GCAAACCTTT GGTGAGAGG TTCAAACAGG AAGGTTTCGA
    351 CTGTTTGGAA AAGCAGGGGT TGCGGCGCAA CGGCCTGTCC GAGCGCGTCC
    401 GATGGTAA

```

This corresponds to the amino acid sequence <SEQ ID 864; ORF131ng-1>:

```

35 1 MEIRVIKYTA TAALFAFTVA GCRLAGWYEC SLSGWCKPR KPAAIDFWDI
    51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
    101 TRDGKPLVER FKQEGFDCLE KQGLRRNGLS ERVRW*

```

ORF131ng-1 and ORF131-1 show 92.6% identity in 135 aa overlap:

```

40 orf131ng-1.pep MEIRVIKYTATAALFAFTVAGCRLAGWYECSSLSGWCKPRKPAAIDFWDIGGESPLSLED
            ||||:|||| |||:||||||| ||:||||||| |||
orf131-1        MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD

orf131ng-1.pep  YEIPLSDGNRSVRANEYESAQKSYFYRKIGKFEACGLDWRTRDGKPLVERFKQEGFDCLE
            ||||| ||||| |||||:||||| ||||| |||||:| ||| |||||
orf131-1        YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQGGFDCLE

45 orf131ng-1.pep KQGLRRNGLSERVRWX
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf131-1        KQGLRRNGLSERVRWX

```

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.



**Example 102**

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 865>

```

1  ATGAAACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGGCGGGCT
51  TGCCGCCATT GCCAAGAAG CGGGGTTTGA AGTCAGCGGT TGCGACGCGA
101 AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
151 TATGAAGGCT TCGATGCCGC TCAGTTGGAC GAATTTAAAG CCGACGTTTA
201 CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
251 TGAACCTCGG CCTGCCCTAT ATTtCGGCC CGCAATGGCT GTCGAAAAC
301 GTGCTGCACC ATCATTGGGT ACTCGGTGTG GCGGGGACgC ACGGCAAAAC
10 351 GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATgCC GGCCTCGCGC
401 CGGGCTTCCT TATtGGCGGC GTACC.GGAA AATtCGGCG TTTCCGCCCG
451 CCTGCCGCAA ACGCCGCGCC AAGACCCGAA CAGCCAATCG CCGTTTTtCG
501 TCATCGAAGC CGACGAATAC GACACCGCCT TTtTCGACAA ACGTTCTAAA
551 TtCGTGCATT ACCGTCCGCG TACCGCCGTG TTGAACAATC TGGAAATTCGA
15 601 CCACGCCGAC ATCTTTGCCG ACTTGGGCGC GATACAGAcC CAGTTCCAAT
651 ACCTCGTGCG TACCGTGCCG TCTGAAGGCT TAATCGTCTG CAACGCGACGG
701 CAGCAAAGCC TGCAAGATAC TTTGGACAAA GGCTGCTGGA CGCCGGTGGG
751 AAAATTCGGC ACGGAACACG GCTGGCA..

```

This corresponds to the amino acid sequence <SEQ ID 866; ORF132>:

```

20 1  MKHIHIIGIG GTFMGGLAAI AKEAGFEVSG CDAKMYPPMS TQLEALGIDV
51  YEGFDAAQLD EFKADVYVIG NVAKRGMDEV EAILNLGLPY ISGPQWLSN
101 VLHHHWLVGV AGTHGKTTTA SMLAWVLEYA GLAPGFLIGG VXGKFRFRFP
151 PAANAAPRPE QPIAVFRHRS RRIRHRLFRQ TFXIRALPSA YRRVEQSGIR
201 PRRHLCRLGR DTDVPLPRA YRAVXRLNRL QRTAAKPARY FGQRLLDAGG
25 251 KIRHGTRLA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 867>:

```

1  ATGAAACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGGCGGGCT
51  TGCCGCCATT GCCAAGAAG CGGGGTTTGA AGTCAGCGGT TGCGACGCGA
101 AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
30 151 TATGAAGGCT TCGATGCCGC TCAGTTGGAC GAATTTAAAG CCGACGTTTA
201 CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
251 TGAACCTCGG CCTGCCCTAT ATTTCCGGCC CGCAATGGCT GTCGAAAAC
301 GTGCTGCACC ATCATTGGGT ACTCGGTGTG GCGGGGACGC ACGGCAAAAC
35 351 GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATGCC GGCCTCGCGC
401 CGGGCTTCCT TATTGGCGGC GTACCGGAAA ATTTCCGGCT TTTCCGCCCG
451 CTGCCGCAA ACGCCGCGCA AGACCCGAAC AGCCAATCGC CGTTTTTCGT
501 CATCGAAGCC GACGAATACG ACACCGCCTT TTTCGACAAA CGTTCTAAAT
551 TCGTGCATTA CCGTCCGCGT ACCGCCGTGT TGAACAATCT GGAATTCGAC
601 CACGCCGACA TCTTTGCCCA CTTGGGCGCG ATACAGACCC AGTTCACATA
40 651 CCTCGTGCGT ACCGTGCCGT CTGAAGGCTT AATCGTCTGC AACGGACGGC
701 AGCAAAGCCT GCAAGATACT TTGGACAAAG GCTGCTGGAC GCCGTGGAA
751 AAATTCGGCA CGGAACACGG CTGGCAGGCC GGCGAAGCCA ATGCCGACGG
801 TCTCGTCGAC GTGTGTCG ACGGCAAAAC CGCCGACGC GTCAAATGGG
851 ATTTGATGGG CAGGCACAAC CGCATGAACG CGCTCGCCGT CATTGCCGCC
45 901 GCGCGTCATG TCGGTGTCGA TATTGAGACC GCCTGCGAAG CCTTGGGCGC
951 GTTTAAAAAC GTCAAACGCC GGATGGAAT CAAAGGCACG GCAAACGCCA
1001 TCACCGTTTA CGACGACTTC GCCCACCACC CGACCGCCAT CGAAACCACG
1051 ATTCAGGTTT TCGGCCAACG CGTCGGCGGC GCGCGCATCC TCGCCGTCCT
1101 CGAACC CGT TCCAACACGA TGAAGCTGGG CACGATGAAG TCCGCCCTGC
50 1151 CTGTAAGCCT CAAAGAAGCC GACCAAGTGT TCTGCTACGC CGGCGGCGTG
1201 GACTGGGACG TCGCCGAAGC CCTCGCGCCT TTGGGCGGCA GGCTGAACGT
1251 CGGCAAAGAC TTCGATGCCT TCGTTGCCGA AATCGTGAAA AACGCCGAAG
1301 TAGGCGACCA TATTTTGGTG ATGAGCAACG GCGGTTTCGG CGGAATACAC
1351 GGAAAGCTGC TGAAGCTTT GAGATAG

```

55 This corresponds to the amino acid sequence <SEQ ID 868; ORF132-1>:

```

1  MKHIHIIGIG GTFMGGLAAI AKEAGFEVSG CDAKMYPPMS TQLEALGIDV
51  YEGFDAAQLD EFKADVYVIG NVAKRGMDEV EAILNLGLPY ISGPQWLSN
101 VLHHHWLVGV AGTHGKTTTA SMLAWVLEYA GLAPGFLIGG VPENFGVSAR
151 LPQTPRQDPN SQSPFFVIEA DEYDTAFFDK RSKFVHYRPR TAVLNNLEFD
60 201 HADIFADLGA IQTQFHYLVR TVPSEGLIVC NGRQOSLQDT LDKGCWTPVE
251 KFGTEHWQQA GEANADGSFD VLLDGKTAGR VKWDLMGRHN RMNALAVIAA
301 ARHVGVDIQT ACEALGAFKN VKRRMEIKGT ANGITVYDDF AHHPTAIETT

```

351 IQGLRQRVGG ARILAVLEPR SNTMKLGTMK SALPVSILKEA DQVFCYAGGV  
 401 DWDVAEALAP LGGRLNVGKD FDAFVAEIVK NAEVGDHILV MSNGGFGGIH  
 451 GKLLEALR\*

Computer analysis of this amino acid sequence gave the following results:

5 Homology with the hypothetical o457 protein of *E.coli* (accession number U14003)

ORF132 and o457 show 58% aa identity in 140 aa overlap:

10 Orf132: 4 IHIIGIGGTFMGGGLAAIAKEAGFEVSGCDAMYPMPSTQLEALGIDVYEGFDAAQLDEFK 63  
 IHI+GI GTFMGGGLA +A++ G EV+G DA +YPPMST LE GI++ +G+DA+QL+ +  
 o457: 3 IHILGICGTFMGGGLAMLARQLGHEVTGSDANVYPPMSTLLEKQGIELIQGYDASQLEP-Q 61  
 Orf132: 64 ADVVYIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWVLGVAGTHGKTTTASML 123  
 D+ +IGN RG VEA+L +PY+SGPQWL + VL WVL VAGTHGKTTA M  
 o457: 62 PDLVIIGNAMTRGNPCVEAVLEKNIPYMSGPQWLHDFVLRDRWVLAVAGTHGKTTTAGMA 121  
 15 Orf132: 124 AWWLEYAGLAPGFLIGGVXG 143  
 W+LE G PGF+IGGV G  
 o457: 122 TWILEQCGYKPGFVIGGVPG 141

Homology with a predicted ORF from *N.meningitidis* (strain A)

20 ORF132 shows 74.6% identity over a 189aa overlap with an ORF (ORF132a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
25	orf132.pep	MKHIHIIGIGGTFMGGGLAAIAKEAGFEVSGCDAMYPMPSTQLEALGIDVYEGFDAAQLD					
	orf132a	MKHIHIIGIGGTFMGGIAAIAKEAGFEVSGCDAMYPMPSTQLEALGIDVYEGFDAAQLD					
		10	20	30	40	50	60
30	orf132.pep	EFKADVYIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWVLGVAGTHGKTTT					
	orf132a	EFKADVYIGNVAKRGMDVVEAILNRLGLPYISGPQWLAENXLHHHWXLGVAXTHGKTTT					
		70	80	90	100	110	120
35	orf132.pep	SMLAWVLEYAGLAPGFLIGGVXGKFR---RFRPPAANAAPRPEQFI-----AVFR					
	orf132a	SMLAWVLEYAGLAPGFXIGGVNPENFSVSARL-PQTPRODPNSQSPFFVIEADEYDTAFED					
		130	140	150	160	170	
40	orf132.pep	HRSRRIRHRLRFQTFXIRALPSAYRVEQSGIRPRRHLCLGRDTPVPLPRAYRAVXRL					
	orf132a	KRSKEFVHYRPRTAVLNNLEFDHADIFADLGAIQTQFHHLVVRTVPSEGLIVCNGRQOSLQD					
		180	190	200	210	220	230

45 The complete length ORF132a nucleotide sequence <SEQ ID 869> is:

1 ATGAAACACA TCCACATTAT CGGTATCGGC GGCACGTTTA TGGGTGGGAT  
 51 TGCCGCCATT GCCAAGAAG CAGGGTTTGA ANTCAGCGGT TGCGATGCGA  
 101 AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG CATAGGCGTG  
 151 TATGAAGGCT TCGACACCGC GCAGTTGGAC GAATTAAAG CCGACGTTTA  
 201 CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGT GAAGCGATTT  
 251 TGAACCGTGG GCTGCCTTAT ATTTCCGGCC CGCAATGGCT GGCTGAAAC  
 301 NTGCTGCACC ATCATTGGNN ACTCGGCGTG GCGGNGACGC ACGGCAAAAC  
 351 GACCACCGCG TCTATGCTCG CGTGGGTTT GGAATATGCC GGAATCGCAC  
 401 CGGGCTTCNT TATCGCGGC GTACCGGAAA ACTTCAGCGT TTCCGCGCGC  
 451 CTGCGCGAAA CGCCGCGCCA AGACCGGAAC AGCCAATCGC CGTTTTTCGT  
 501 CATTGAAGCC GACGAATACG ACACCGCGTT TTTCGACAAA CGCTCCAAAT  
 551 TCGTGCATTA CCGTCCGCGT ACCGCGGTGT TGAACAATCT GGAATTCGAC  
 601 CACGCCGACA TCTTCGCCGA TTGGGCGCG ATACAGACCC AGTTCCACCA  
 651 CCTCGTGCGT ACCGTCCCGT CTGAAGGCCT CATCGTCTGC AACGGACGGC  
 701 AGCAAAGCCT GCAAGACACT TTGGACAAAG GCTGCTGGAC GCCGGTGAAA  
 751 AAATTCGGCA CGGAACACGG CTGGCAGGCC GGCGAAGCCA ATGCCGATGG

**This encodes a protein having amino acid sequence <SEO ID 870>:**

15

20

ORF132a and ORF132-1 show 93.9% identity in 458 aa overlap:

25

```

orf132a.pep      MKHHHHIIGGGTFMGGIAAIAKEAGFEXSGCDKMYPPMSTQLEALGIGVYEGFDTAQLD
                  |||||||||||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf132-1         MKHHHHIIGGGTFMGGGLAAIAKEAGFEVSGCDKMYPPMSTOLEALGIDVYEGEDAAOLD

```

30

```
orf132a.pep      EFKADVYVIGNVAKRGMDVVEAILNRGLPYISGPQWLAENXLHHHWXLGVAXTHGKTTTA
                  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
orf132-1         EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPWLSENVLHHHWLGVAGTHGKTTTA
```

35

orf132a.pep SMLAWVLEYAGLAPGFXIGGVNPENFSVSARLPQTTPRQDPNSQSPPFFVIEADEYDTAFFDK  
| | | | | : | | | | |  
orf132-1 SMLAWVLEYAGLAPGLIGGVNPFNGVSARLPOTPRDPSNSQSPPFFVIEADEYDTAFFDK

40

```

orf132a.pep      RSKFVHYRPRTAVLNLEFDHADIFADLGAIQTQFHHLVTRVPSEGLIVCNGRQOQLQDT
                  |||||
orf132-1         RSKFVHYRPRTAVLNLEFDHADIFADLGAIQTQFHYLVTRVPSEGLIVCNGROOQLQDT

```

45

orf132a.pep LDKGCWTPVEKFGTEHGWOAGEANADGSFDVLLDGKKAGHVAWSLMGGHNRMNALAVIAA  
||| | | | | | | | | | | | | | | | | | | | | | | | | | | : | : ||  
orf132-1 LDKGCWTPVEKFGTEHGWOAGEANADGSFDVLLDGKTAGRVKWDLMGRHNRMNALAVIAA

50

```

orf132a.pep      ARHAGVDIQTACEALSTFKNVKRRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf132-1         ARHVGVDIQTACEALGAFKNVKRRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG

orf132a.pep      ARILAVLEPRSNMTKLGTMKAALPASLKEADQVFXYAGGADWDVAEALAPLGGRLHVGKD
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf132-1         ARILAVLEPRSNMTKLGTMKSALPVSILKEADQVFCYAGGVWDVAEALAPLGGRLNVGKD

```

55

```

orf132a.pep      FDFAEIVKNAEAGDHILVMSNGGFGGIHTKLLDALRX
                  |||||:|||||:|||||:|||||
orf132-1         FDFAEIVKNAEVDHILVMSNGGFGGIHGKLLAEALRX

```

ORF132 shows 89.6% identity over 259 aa overlap with a predicted ORF (ORF132ng) from *N. gonorrhoeae*:

60

<b>orf132.pep</b>	MKHIHIIIGIGGTFMGGLAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLD	60
	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
<b>orf132ng</b>	MKHIHIIIGIGGTFMGGIAAIAKEAGFKVSGCDAKMYPPMSTOLEALGIGVHEGFDAAOLE	60

-471-

	orf132.pep	EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTA	120
	orf132ng	EFQADIYVIGNVARRGMDVVEAILNRGLPYISGPQWLAENVLHHHWLVGVAGTHGKTTTA	120
5	orf132.pep	SMLAWVLEYAGLAPGFLIGGVXGKFRFRPPAANAAPRPEQPIAVFRHRSRRIRHRLFRQ	180
	orf132ng	SMLAWVLEYAGLAPGFLIGGVPGKFRFRPPPTANAASRPEQQIAVFRHRSRRIRHRLFRQ	180
10	orf132.pep	TFXIRALPSAYRRVEQSGIRPRRHLCRLGRDTPVPLPRAYRAVXRLNRLQRTAAKPARY	240
	orf132ng	TLQIRALSPAYRRVEQSGIRPRRHLLRLGRDTPVPPRAHRTIRRHRLQRTAAKPARY	240
	orf132.pep	FGQRLLDAGGKIRHGTRLA	259
15	orf132ng	FGQRLLDAGGKIRHRTLADW	261

An ORF132ng nucleotide sequence <SEQ ID 871> was predicted to encode a protein having amino acid sequence <SEQ ID 872>:

	1	MKHIHIIGIG	GTFMGGIAAI	AKEAGFKVSG	CDAKMYPPMS	TQLEALGIGV
20	51	HEGFDAQLE	EFQADIYVIG	NVARRGMDV	EAILNRGLPY	ISGPQWLAEN
	101	VLHHHWLVGV	AGTHGKTTTA	SMLAWVLEYA	GLAPGFLIGG	VPKGFRFRP
	151	PTANAASRPE	QQIAVFRHRS	RRIRHRLFRQ	TLQIRALSPA	YRRVEQSGIR
	201	PRRHLRRLGR	DTDPVPPRA	HRTIRRHRL	QRTAAKPARY	FGQRLLDAGG
	251	KIRHRTLAD	W*			

Further work revealed the following gonococcal DNA sequence <SEQ ID 873>:

25	1	ATGAAACACA	TCCACATTAT	CGGTATCGGC	GGCACGTTTA	TGGGCGGGAT
	51	TGCCGCCATT	GCCAAAGAAG	CCGGGTTCAA	AGTCAGCGGT	TGCGACGCGA
	101	AGATGTATCC	GCCGATGAGC	ACCCAGCTCG	AAGCCTTGGG	CATAGGCGTA
	151	CACGAAGGCT	TCGATGCCGC	GCAGTTGGAA	GAATTTCAAG	CCGATATTTA
30	201	CGTCATCGGC	AATGTCGCCA	GGCGCGGGAT	GGATGTGGTC	GAGGCGATTT
	251	TGAACCGTGG	GCTGCCTTAT	ATTTCCGGCC	CGCAATGGCT	GGCTGAAAac
	301	GTGCTgcacc	atcaTTGGgt	ACTCGGCGTG	GcagggacGC	ACGGcaaAac
	351	gaccaCcGcg	tCCATGCTCG	CCTGGGTCTT	GGAATATGCC	GGACTCGCGC
	401	CGGGCTTCCT	CATCGGCGGt	gtaccggaAA	ATTTCCGGCGT	TTCCGCCCGC
	451	CTACCGCAAA	CGCCGCGTCA	AGACCGGAAC	AGCAAATCGC	CGTTTTTCGT
35	501	CATCGAAGCC	GACGAATACG	ACACCGCCTT	TTTCGACAAA	CGCTCCAAAT
	551	TCGTGCATTA	TCGCCCGCGT	ACCGCCGTGT	TGAACAATCT	GGAATTCGAC
	601	CACGCCGACA	TCTTCGCCGA	CTTGGGCGCG	ATACAGACCC	AGTTCACCA
	651	CCTCGTGC	ACCGTACCAT	CCGAAGGCCT	CATCGTCTGC	AACGGACAGC
40	701	AGCAAAGCCT	GCAAGATACT	TTGGACAAAG	GCTGCTGGAC	GCCGGTGGAA
	751	AAATTCGCGA	CCGGACACGG	CTGGCAGATT	GGTGAAGTCA	ATGCCGACGG
	801	CTCGTTCGAC	GTATTGCTTG	ACGGCAAAAA	AGCCGGACAC	GTCCGATGGG
	851	ATTTGATGGG	CGGACACAAC	CGCATGAACG	CGCTCGCCGT	CATCGTGGC
	901	GCACGCCATG	CCGGAGTCGA	TGTTCAAGACG	GCCTGCGAAG	CCTTGGGTGC
	951	GTTTAAAAAC	GTCAAACGCC	GCATGGAAT	CAAAGGCACG	GCAAACGGCA
45	1001	TCACCGTTTA	CGACGATTTC	GCCCACCACC	CGACCGCCAT	CGAAACCACG
	1051	ATTCAAGGTT	TGCGCCAACG	TGTCGGCGGC	GCGCGCATCC	TGCGCCGTCT
	1101	CGAGCCGCGT	TCCAACACCA	TGAAACTCGG	CACGATGAAG	TCCGCCCTGC
	1151	CCGCAAGCCT	CAAAGAAGCC	GACCAAGTGT	TCTGTACGCG	CGGCGGCGCG
	1201	GACTGGGACG	TTGCCGAAGC	CCTCGCGCCT	TTGGGCTGCA	GGCTGCGCGT
50	1251	CGGTAAAGAT	TTGATACCT	TCGTTGCCGA	AATTGTGAAA	AACGCCCGAA
	1301	CCGGCGACCA	TATTTTGGTG	ATGAGCAACG	GCGGTTTCGG	CGGAATACAC
	1351	ACCAAACGTC	TGGACGCTTT	GAGATAG		

This corresponds to the amino acid sequence <SEQ ID 874; ORF132ng-1>:

	1	MKHIHIIGIG	GTFMGGIAAI	AKEAGFKVSG	CDAKMYPPMS	TQLEALGIGV
55	51	HEGFDAQLE	EFQADIYVIG	NVARRGMDV	EAILNRGLPY	ISGPQWLAEN
	101	VLHHHWLVGV	AGTHGKTTTA	SMLAWVLEYA	GLAPGFLIGG	VPENFGVSAR
	151	LPQTPRQDPN	SKSPFFVIEA	DEYDTAFFDK	RSKFVHYRPR	TAVLNNLEFD
	201	HADIFADLGA	IQTQFHHLVR	TVPSEGLIVC	NGQQQSLQDT	LDKGCWTPVE
	251	KFGTGHGWQI	GEVNADGSFD	VLLDGKKAGH	VAWDLMGGHN	RMNALAVIAA
60	301	ARHAGVDVQT	ACEALGAFKN	VKRRMEIKGT	ANGITVYDDF	AHHPTAIETT
	351	IQGLRQRVGG	ARILAVLEPR	SNTMKLGTMK	SALPASLKEA	DQVFCYAGGA
	401	DWDVAEALAP	LGCLRLRVGKD	FDTFVAEIVK	NARTGDHILV	MSGNGFGGIH
	451	TKLLDALR*				

ORF132ng-1 and ORF132-1 show 93.2% identity in 458 aa overlap:

```

5  orf132ng-1.pep MKHIHIIGIGGTFMGGIAAAIAKEAGFKVSGCDAKMYPPMSTQLEALGIGVHEGFDAQAQLE
   orf132-1      MKHIHIIGIGGTFMGGIAAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAQAQLD
10  orf132ng-1.pep EFQADIYVIGNVARRGMDVVEAILNRGLPYISGPQWLAENVLHHHWLVGVAGTHGKTTTA
   orf132-1      EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTA
15  orf132ng-1.pep SMLAWVLEYAGLAPGFLIGGVPENFVGSARLPQTPRQDPNSKSPFFVIEADEYDTAFFDK
   orf132-1      SMLAWVLEYAGLAPGFLIGGVPENFVGSARLPQTPRQDPNSQSPFFVIEADEYDTAFFDK
20  orf132ng-1.pep RSKFVHYRPRTAVLNNLEFDHADIFADLGAIQTQFHHLVTRVPSEGLIVCNGQQQSLQDT
   orf132-1      RSKFVHYRPRTAVLNNLEFDHADIFADLGAIQTQFHYLVTRVPSEGLIVCNGRQQSLQDT
25  orf132ng-1.pep LDKGCWTPVEKFGTGHGWQIGEVNADGSFDVLLDGKKAGHVAWDLMGHNRMNALAVIAA
   orf132-1      LDKGCWTPVEKFGTEHGQAGEANADGSFDVLLDGKTAGRVKWDLMGRHNRMNALAVIAA
30  orf132ng-1.pep ARHAGVDVQTACEALGAFKNVKRRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG
   orf132-1      ARHVGVDIQTACEALGAFKNVKRRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG
35  orf132ng-1.pep ARILAVLEPRSNMTKLGTMKSALPASLKEADQVFCYAGGADWDVAEALAPLGCRRLRVGKD
   orf132-1      ARILAVLEPRSNMTKLGTMKSALPVSLKEADQVFCYAGGVDWDVAEALAPLGGRRLNVGKD
40  orf132ng-1.pep FDTFVAEIVKNARTGDHILVMSNGGFGGIHTKLLDALRX
   orf132-1      FDAFVAEIVKNAEVDHILVMSNGGFGGIHGKLEALRX

```

In addition, ORF132ng-1 is homologous to a hypothetical *E.coli* protein:

```

35  pir||S56459 hypothetical protein o457 - Escherichia coli >gi|537075 (U14003)
   ORF_o457 [Escherichia coli] >gi|1790680 (AE000494) hypothetical 48.5 kD protein
   in fbp-pmba intergenic region [Escherichia coli] Length = 457
   Score = 474 bits (1207), Expect = e-133
   Identities = 249/439 (56%), Positives = 294/439 (66%), Gaps = 13/439 (2%)
40  Query: 22  KEAGFKVSGCDAKMYPPMSTQLEALGIGVHEGFDAQAQLEEFQADIYVIGNVARRGMDVVE 81
   Sbjct: 21  RQLGHEVTGSDANVYPPMSTLLEKQGIELIQGYDASQLEP-QPDLVIIGNAMTRGNPCVE 79
45  Query: 82  AILNRGLPYISGPQWLAENVLHHHWLVGVAGTHGKTTTASMLAWVLEYAGLAPGFLIGGV 141
   Sbjct: 80  AVLEKNIPYMSGPQWLHDFVLRDRWVLAVAGTHGKTTTAGMATWILEQCQGYKPGFVIGGV 139
50  Query: 142  PENFGVSARLPQTPRQDPNSKSPFFVIEADEYDTAFFDKRSKFVHYRPRTAVLNNLEFDH 201
   Sbjct: 140  PGNFEVSAHL-----GESDFFVIEADEYDCAFFDKRSKFVHYCPRTLILNNLEFDH 190
55  Query: 202  ADIFADLGAIQTQFHHLVTRVPSEGLIVCNGQQQSLQDTLDKGCWTPVEKFGTGHGWQIG 261
   Sbjct: 191  ADIFDDLKAIQKQFHHLVRIVPGQGRIIWPENDINLKQTMAMGCWSEQELVGEQGHWOAK 250
60  Query: 262  EVNADGS-FDVLLDGKKAGHVAWDLMGHNRMNALAVIAAARHAGVDVQTACEALGAFKN 320
   Sbjct: 251  KLTTDASEWEVLLDGEKVGEVKWSLVGEHNMHNLMAIAAARHVGVPADAANALGSFIN 310
65  Query: 321  VKRRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG-ARILAVLEPRSNMTKLGTM 379
   Sbjct: 311  ARRRLELRGEANGVTVYDDFAHHPTAILATLAALRGKVGGTARIIVLEPRSNMTKMGIC 370
70  Query: 380  KSALPASLKEADQVF-CYAGGADWDVAEALAPLGCRRLRVGKDFDTFVAEIVKNARTGDHI 438
   Sbjct: 371  KDDLAPSLGRADEVFLQPAHIPWQVAEVAEACVQPAHWSGDVDTLADMVVKTAQPGDHI 430
75  Query: 439  LVMSNGGFGGIHTKLLDAL 457

```

LVMSNGGFGGIH KLLD L  
 Sbjct: 431 LVMSNGGFGGIHQKLLDGL 449

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 ORF132-1 (26.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 20A shows the results of affinity purification of the His-fusion protein, and Figure 20B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 20C) and ELISA (positive result). These  
 10 experiments confirm that ORF132 is a surface-exposed protein, and that it is a useful immunogen.

### Example 103

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 875>

```

1  ..CCGGGCTATT ACGGCTCGGA TGACGAATTT AAGCGGGCAT TCGGAGAAAA
15  51  CTCGCCGACA TmCAAGAAAC ATTGCAACCG GAGCTGCGGG ATTTATGAAC
    101 CCGTATTGAA AAAATACGGC AAAAAGCGCG CCAACAACCA TTCGGTCAGC
    151 ATTAGTGCGG ACTTCGGCGA TTATTTTCATG CCGTTCGCCA GCTATTCGCG
    201 CACACACCGT ATGCCCAACA TCCAAGAAAT GTATTTTTC CAAATCGGCG
    251 ACTCCGGCGT TCACACCGCC TTAAAACCAG AGCGCGCAA CACTTGGCAA
    301 TTTGGCTTCr ATACCTATAA AAAAGGATTG TTAACAACAG ATGATACATT
20  351 AGGATTAAAA CTGGTCGGCT ACCGCAGCCG CATCGACAAC TACATCCACA
    401 ACGTTTACGG GAAATGGTGG GATTTGAACG GGGATATTC GAGCTGGGTC
    451 AGCAGCACCG GGCTTGCCTA CACCATCCAA CATCGCrATT TCaWAGACAA
    501 AGTGCATCAA nnnnnnnnnn nnnnnnnnnn nnnnTACGAT TATGGGCGTT
    551 TTTTCACCAA CCTTTCTTAC GCCTATCAAA AAAGCACGCA ACCGACCAAC
25  601 TTCAGCGATG CGAGCGAATC GCCCAACAAT GCGTCCAAAG AAGACCAACT
    651 CAAACAAGGT TATGGGTGA GCAGGGTTTC CGCCCTGCCG CGAGATTACG
    701 GACGTTTGA AGTCGGTACG CGCTGGTTGG GCAACAACT GACTTTGGGC
    751 GCGCGCATGC GCTATTTCCG CAAGAGCATC CGCGCGACGG CTGAAGAACG
    801 CTATATCGAC GGCACCAACG GGGGAAATAC CAGCAATTTC CGGCAACTGG
30  851 GCAAGCGTTC CATCAACAA ACCGAAACTC TTGCCCGCCA GCCTTTGATT
    901 TTWgATTTTa ACGCCGCTTA CGAGCCGAAG AAAAACCTTA TTTTCCGCGC
    951 CGAAGTCAA AATCTGTTCG ACAGGCGTTA TATCGATCCG CTCGATGCGG
    1001 GCAATGATGC GGCAAC.GAG CGTTATTACA GCTCGTTCTGA CCCGAAAGAC
    1051 AAGGACrAG ACGTAACGTG TAATGCTGAT AAAACGTTGT GCaACGGCAA
35  1101 ATACGGCGGC ACAAGCAAAA GCGTATTGAC CAATTTTGCA CGCGGACGCA
    1151 CCTTTTtGAT GACGATGAGC TACAAGTTTT AA
  
```

This corresponds to the amino acid sequence <SEQ ID 876; ORF133>:

```

1  ..PGYYGSDDEF KRAFGENSPT XKKHCNRSCG IYEPVLKKYG KKRANNHSVS
40  51  ISADFGDYFM PFASYSRTHR MPNIQEMYFS QIGDSGVHTA LKPERANTWQ
    101 FGFXTYKGL LKQDDTLGLK LVGYRSRIDN YIHNVYKWW DLNGDIPSWV
    151 SSTGLAYTIQ HRXFXDKVHQ XXXXXXXXYD YGRFFTNLSY AYQKSTQPTN
    201 FSDASESPNN ASKEDQLKQG YGLSRVSALP RDYGRLEVGT RWLGNKLTIG
    251 GAMRYFGKSI RATAEERYID GTNGGNTSNF RQLGKRSIKQ TETLARQPLI
    301 XDFNAAYEPK KNLIrFAEVK NLFDRRYIDP LDAGNDAAXE RYYSSFDPKD
45  351 KDXDVTCTNAD KTLcNGKYGG TSKSVLTNFA RGRTFLMTMS YKF*
  
```

Further work revealed the further partial DNA sequence <SEQ ID 877>:

```

1  GAGGCGCAGA TACAGGTTTT GGAAGATGTG CACGTCAAGG CGAAGCGCGT
51  ACCGAAAGAC AAAAAAGTGT TTACCGATGC GCGTGCCGTA TCGACCCGTC
101 AGGATATATT CAAATCCAGC GAAAACCTCG ACAACATCGT ACGCAGCATC
50  151 CCCGGTGCGT TTACACAGCA AGATAAAAGC TCGGGCATTG TGTCTTTGAA
    201 TATTCGCGGC GACAGCGGGT TCGGGCGGGT CAATACGATG GTGGACGGCA
    251 TCACGCAGAC CTTTTATTTC ACTTCTACCG ATCGGGGCAG GGCAGGCGGT
  
```

-474-

301 TCATCTCAAT TCGGTGCATC TGTCGACAGC AATTTTATTG CCGGACTGGA  
 351 TGTCGTCAAA GGCAGCTTCA GCGGCTCGGC AGGCATCAAC AGCCTTGCCG  
 401 GTTCGGCGAA TCTGCGGACT TTAGGCGTGG ATGACGTCGT TCAGGGCAAT  
 451 AATACCTACG GCCTGCTGCT AAAAGGTCTG ACCGGCACCA ATTCAACCAA  
 501 AGGTAATGCG ATGGCGGCGA TAGGTGCGCG CAAATGGCTG GAAAGCGGAG  
 551 CATCTGTCGG TGTGCTTTAC GGGCACAGCA GGCGCAGCGT GGCACAAAT  
 601 TACCGCGTGG GCGCGGCGG GCGAGCACATC GGAATTTTGG GCGCGGAATA  
 651 TTTGGAACGG CGCAAGCAGC GATATTTTGT ACAAGAGGGT GCTTTGAAAT  
 701 TCAATTCCGA CAGCGGAAAA TGGGAGCGGG ATTTACAAAG GCAACAGTGG  
 751 AAATACAAGC CGTATAAAAA TTACAACAAC CAAGAACTAC AaAAATACAT  
 801 CGAAGAGCAT GACAAAAGCT GCGGGGAAAA CCTg. CaCCG CAATACGACA  
 851 TTACCCCAT CGATCCGTCC AGCCTGAAGC AGCAGTCGGC AGGCAATCTG  
 901 TTAAATTGG AATACGACGG CGTATTCAAT AAATACACGG CGCAATTTCG  
 951 CGATTTAAAC ACCAAAATCG GCAGCCGCAA AATCATCAAC CGCAATTATC  
 1001 AGTTCAATTA CGGTTTGTCT TTGAACCCGT ATACCAACCT CAATCTGACC  
 1051 GCAGCCTACA ATTCGGGCAG GCGAAAATAT CCGAAAGGGT CGAAGTTTAC  
 1101 AGGCTGGGGG CTTTAAAGG ATTTTGAAAC CTACAACAAC GCGAAAATCC  
 1151 TCGACCTCAA CAACCCGCC ACCTTCGGC TGCCCGCGCA AACCGAGTTG  
 1201 CAAACCACTT TGGGCTTCAA TTATTTCCAC AACGAATACG GCAAAAACCG  
 1251 CTTTCCTGAA GAATTGGGGC TGTTTTTCGA CGGTCTGAT CAGGACAACG  
 1301 GGCTTTATTC CTATTTGGGG CGGTTTAAGG GCGATAAAGG GCTGCTGCC  
 1351 CAAAAATCAA CCATTGTCCA ACCGGCCGGC AGCCAATATT TCAACACGTT  
 1401 CTAATTCGAT GCCGCGCTCA AAAAAGACAT TTACCGCTTA AACTACAGCA  
 1451 CCAATACCGT CCGCTACCGT TTCGGCGGCG AATATACGGG CTATTACGGC  
 1501 TCGGATGACG AATTTAAGCG GGCATTTCGA GAAACTCGC CGACATACAA  
 1551 GAAACATTGC AACCGGAGCT GCGGGATTTA TGAACCCGTA TTGAAAAAT  
 1601 ACGGCAAAAA GCGCGCCAAC AACCATTTCG TCAGCATTAG TCGCGACTTC  
 1651 GCGGATTATT TCATGCCGTT CGCCAGCTAT TCGCGCACAC ACCGTATGCC  
 1701 CAACATCCAA GAAATGTATT TTTCCCAAAT CCGCGACTCC GCGGTTTACA  
 1751 CCGCCTTAAA ACCAGAGCGC GCAAACACTT GGCAATTTGG CTTCATATCC  
 1801 TATAAAAAAG GATTGTTAAA ACAAGATGAT ACATTAGGAT TAAAAGTGGT  
 1851 CGGCTACCGC AGCCGCATCG ACAACTACAT CCACAACGTT TACGGGAAAT  
 1901 GGTGGGATTT GAACGGGGAT ATTCCGAGCT GGGTCAGCAG CACCGGGCTT  
 1951 GCCTACACCA TCCAACATCG CAATTTCAA GACAAAGTGC ACAAACACGG  
 2001 TTTTGAGTTG GAGCTGAATT ACGATTATGG GCGTTTTTTC ACCAACCTTT  
 2051 CTTACGCCTA TCAAAAAAGC ACGCAACCGA CCAACTTCAG CGATGCGAGC  
 2101 GAATCGCCCA ACAATGCGTC CAAAGAAGAC CAACTCAAAC AAGGTTATGG  
 2151 GTTGAGCAGG GTTTCGCCCC TGCCGCGAGA TTACGGACGT TTGGAAGTCG  
 2201 GTACGCGCTG GTTGGGCAAC AAAGTACTT TGGGCGGCGC GATGCGCTAT  
 2251 TCCGGCAAGA GCATCCGCGC GACGGCTGAA GAACGCTATA TCGACGGCAC  
 2301 CAACGGGGGA AATACCAGCA ATTTCCGGCA ACTGGGCAAG CGTTCCATCA  
 2351 AACAAACCGA AACTCTTGCC CGCCAGCCTT TGATTTTGA TTTTACGCC  
 2401 GCTTACGAGC CGAAGAAAAA CCTTATTTTC CGCGCCGAAG TCAAAAATCT  
 2451 GTTCGACAGG CGTTATATCG ATCCGCTCGA TGCGGGCAAT GATGCGGCAA  
 2501 CGCAGCGTTA TTACAGCTCG TTCGACCCGA AAGACAAGGA CGAAGACGTA  
 2551 ACGTGTAATG CTGATAAAAC GTTGTGCAAC GGCAATACG GCGGCACAAG  
 2601 CAAAAGCGTA TTGACCAATT TTGCACGCGG ACGCACCTTT TTGATGACGA  
 2651 TGAGCTACAA GTTTTAA

This corresponds to the amino acid sequence <SEQ ID 878; ORF133-1>:

50 1 EAQIQVLEDV HVKAKRVPKD KKVFTDARAV STRQDIFKSS ENLDNIVRSI  
 51 PGAFTQDDKS SGIVSLNIRG DSGFGRVNTM VDGITQTFYS TSTDAGRAGG  
 101 SSQFGASVDS NFIAGLDVVK GSFSGSAGIN SLAGSANLRT LGVDDVVOGN  
 151 NTYGLLLKGL TGTNSTKGNA MAAIGARKWL ESGASVGVLY GHSRRSVAQN  
 201 YRVGGGQHI GNFGAEYLER RKQRYFVOEG ALKFNSDSGK WERDLQRQQW  
 55 251 KKYKYNNYN QELQYIEEH DKSWRENLP QYDITPIDPS SLKQQSAGNL  
 301 FKLEYDGVFN KYTAQFRDLN TKIGSRKIIN RNYQFNGLS LNPYTNLNL  
 351 AAYNSGRQKY PKGSKFTGWG LLKDFETYN AKILDNLNTA TFRLPRETEL  
 401 QTTLGFNYFH NEYGKNRFE ELGLFFDGPD QDNGLYSYLG RFKGDGKLLP  
 451 QKSTIVQPAG SQYFNTFYFD AALKKDIYRL NYSTNTVGYR FGGEYTGYYG  
 60 501 SDDEFKRAF ENSPTYKKHC NRSCGIYEPV LKKYGKKRAN NHSVVISADP  
 551 GDYFMFFASY SRTHMPNIQ EMYFSQIGDS GVHTALKPER ANTWQFQENT  
 601 YKKGLLKQDD TLGLKLVGYS SRIDNYIHN YGKWWDLNGD IPSWVSSTGL  
 651 AYTIOHRNFK DKVHKHGFEL ELNYDYGRFF TNLSYAYQKS TQPTNFSAS  
 701 ESPNNASKED QLKQGYGLSR VSALPRDYGR LEVGTRWLGN KLTLLGAMRY  
 751 FGKSIIRATAE ERYIDGTNGG NTSNFRQLGK RSIKQTETLA RQPLIFDFYA  
 801 AYEPKKNLIF RAEVKNLFDR RYIDPLDAGN DAATQRYYS FDKPKDKEDV  
 851 TCNADKTLN GYGGTSKSV LTNFARGRTF LMTMSYKF\*

Computer analysis of this amino acid sequence gave the following results:

Homology with with the probable TonB-dependent receptor HI121 of *H. influenzae* (accession number U32801)

ORF133 and HI121 show 57% aa identity in 363aa overlap:

```

5  Orf133: 31 IYEPVLKKYGKKRANNHSVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTA 90
    I EP+L K G K+A NHS ++SA+ DYFMPF +YSRTHRMPNIQEM+FSQ+ ++GV+TA
    HI121: 563 INEPILHKSGHKKAFNHSATLSAELSDYFMPFFTSRTHRMPNIQEMFFSQVSNAQVNTA 622

10 Orf133: 91 LKPERANTWQFGFXTYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWV 150
    LKPE+++T+Q GF TYKKGL QDD LG+KLVGYRS I NYIHNVYG WW +P+W
    HI121: 623 LKPEQSDTYQLGFNTYKKGLFTQDDVLGVKLVGYRSFIKNIHNVYGVWW--RDGMPTWA 680

15 Orf133: 151 SSTGLAYTIQHRXFXDKVHXXXXXXXXXXDYGRFFTNLSYAYQKSTQPTNFSASESPNN 210
    S G YTI H+ + V YD GRFF N+SYAYQ++ QPTN++DAS PNN
    HI121: 681 ESNQFKYTIHQNYKPIVKKSGVELEINYDMGRFFANVSAYQRTNQPTNYADASPRPNN 740

20 Orf133: 211 ASKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKLTGGLGAMRYFGKSIRATAEERYID 270
    AS+ED LKQGYGLSRVS LP+DYGRLE+GTRW KLTGL A RY+GKS RAT EE YI+
    HI121: 741 ASQEDILKQGYGLSRVSMLPKDYGRLELGTWFDQKLTGLAARYYGKSKRATIEEYIN 800

25 Orf133: 271 GTNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLIFRAEVKNLFDRIYIDP 330
    G+ + R+ ++K+TE + +QP+I D + +YEP K+LI +AEV+NL D+RY+DP
    HI121: 801 GSR-FKKNLTLRRENYAVKKTEDIKKQPIILDHVSYPEIKDLIIKAEVQNLLDKRYVDP 859

    Orf133: 331 LDAGNDAAXERYSSFDPKDKDXDVTNADKTLNCGKYGGTSKSVLTNFARGRTFLMTMS 390
    LDAGNDAA +RYSS + + C D + C GG+ K+VL NFARGRT+++++
    HI121: 860 LDAGNDAASQRYSSL-----NNSIECAQDSSAC----GGSDKTVLYNFARGRTYILSLN 910

    Orf133: 391 YKF 393
    YKF
    HI121: 911 YKF 913
  
```

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF133 shows 90.8% identity over a 392aa overlap with an ORF (ORF133a) from strain A of *N. meningitidis*:

```

35  orf133.pep          10      20      30
    PGGYGSDDDEFKRAFGENSPYKXKHCNRSCTGI
    orf133a      FYFDAALKKDIYRLNYSNTNTVGYRFGGXYTGYYXSDDEFKRAFGENSPYKXKHCNQSCGI
    450      460      470      480      490      500

40  orf133.pep          40      50      60      70      80      90
    YEPVLKKYGKKRANNHSVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL
    orf133a      YEPVLKKYGKKRANNHSVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL
    510      520      530      540      550      560

45  orf133.pep          100     110     120     130     140     150
    KPERANTWQFGFXTYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVS
    orf133a      KPERANTWQFGFNTYKKGLLKQDDTLGLKLVGYRSRIDXYIHNVYGKWWDLNGNIPSWVS
    570      580      590      600      610      620

50  orf133.pep          160     170     180     190     200     210
    STGLAYTIQHRXFXDKVHXXXXXXXXDYGRFFTNLSYAYQKSTQPTNFSASESPNNA
    orf133a      STGLAYTIQHRNFKDKVHKHGFELNLDYXRFFTNLSYAYQKSTQPTNFSASESPNNA
    630      640      650      660      670      680

55  orf133.pep          220     230     240     250     260     270
    SKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKLTGGLGAMRYFGKSIRATAEERYIDG
    orf133a      SKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKLTGGLGAMRYFGKSIRATAEERYIDX
    690      700      710      720      730      740

60  orf133.pep          280     290     300     310     320     330
    TNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLIFRAEVKNLFDRIYIDPL
    65  orf133a
  
```



```

||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
orf133a      TNGXXTSNFRQLGKRSIXQTETLARQPLIFDXYYAAYEPKKXLI FRAEVKNLFDRRYIDPL
750          760          770          780          790          800

           340         350         360         370         380         390
orf133.pep   DAGNDAAAXERYSSFPDPKD KDXDVT CNADKTLCNGKYGGT SKSVLTNFARGRTFLMTMSY
            ::::: ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
orf133a     DAGNDAATQRYSSFPDPKD DEEVC NDNTLCNGKYGGT SKSVLTNFARGXTFLITMSY
180          190          200          210          220          230          240          250          260          270
orf133.pep   KFX
            |||
orf133a     KFX
            270

```

A partial ORF133a nucleotide sequence <SEQ ID 879> is:

	1	AAAGACAAAA	AAGTGTTTAC	CGATGCGCGT	GCCGTATCGA	CCCGTCAGGA
	51	TATATTCAAA	TCCANCGAAA	ACCTCGACAA	CATCGTACGC	ANCATCCCCG
20	101	GTGCGTTTAC	ACANCAANAT	AAAAGCTCGG	GCNTTGTGTC	TTTGAATATT
	151	CGCNGCGACA	CGGGGTTCGG	GCGGGTCAAT	ACNATGGTNG	ACGGCATCAC
	201	NCANACCTTT	TATTCGACTT	CTACCGATGC	GCGCAGGGCA	GGCGGTTTCA
	251	CTCAATTCGG	TGCATCTGTC	GACAGCAATT	TTATNGCCGG	ACTGGATGTC
	301	GTCAAAGGCA	GCTTCAGCGG	CTCGGCAGCG	ATCAACAGCC	TTGCCGGTTC
25	351	GCGCAATCTG	CGGACTTTAN	GCGTGGATGA	TGTCGTTTCA	GGCAATANTA
	401	CNTACGGCCT	GCTGCTAAAA	GGTCTGACCG	GTCCAAATTC	AACCAAGGTT
	451	AATGCGATGG	CGGCGATAGG	TGCGCGCAAA	TGGCTGGAAA	GCGGAGCATC
	501	TGTCGGTGTG	CTTTACGGGC	ACAGCAGGCG	CAGCGTGGCG	CAAAATTACC
	551	GCGTGGGCGG	CGCGGGCGAG	CACATCGGAA	ATTTTGGCGC	GGAAATATCTG
30	601	GAACGACGCA	GAAACAGATA	TTTTGAGCAA	GAAGCGGGT	TGAAATTTCAA
	651	TTCCAACAGC	GGAAAATGGG	AGCGGGATTT	CCAAAAGTCG	TACTGGAAAA
	701	CCAAGTGTA	TCAAAAATAC	GATGCCCCCC	AAGAACTGCA	AAAATACATC
	751	GAAGGTCATG	ATAAAAGCTG	GCGGGAAAAAC	CTGGCGCCGC	AATACGACAT
	801	CACCCCCATC	GATCCGTCCA	GCCTGAAGCN	GCAGTCGGCA	GGCAACCTGT
35	851	TTAAATTGGA	ATACGACGGC	GTATTCAATA	ATAACACGGC	GCAATTTTCG
	901	GATTTAAACA	CCAAAATCGG	CAGCCGCAAA	ATCATCAACC	GCAATTATCA
	951	ATTCAATTAC	GGTTTGTCTT	TGAACCCGTA	TACCAACCTC	AATCTGACCG
	1001	GACGCTACAA	TTCCGGGCAGG	CAGAAATATC	CGAAAGGGTC	GAAGTTTACA
	1051	GGCTGGGGGC	TTTTNAAAAG	TTTTGAAACC	TACAACAACG	CAAAAATCCT
40	1101	CGACCTCANC	AACACCTCCA	CCTTCCGGCT	GCCCCGTGAA	ACCGAGTTGC
	1151	AAACCACTTT	GGGCTTCAAT	TATTTCCACA	ACGAATACGG	CAAAAACCCG
	1201	TTTCCTGAAG	AATTGGGGCT	GTTTTTCGAG	GGTCCGGATC	ANGACAACGG
	1251	GCTTTATTCC	TATTTGGGGC	GGTTTAAGGC	CGATAAAGGG	TCGCTGCCCC
	1301	AAAAATCAAC	CATTGTCCAA	CCGGCCGGCA	GCCAATATTT	CAACACGTTT
45	1351	TACTTCGATG	CCGCGCTCAA	AAAAGACATT	TACCGCTTAA	ACTACAGCAC
	1401	CAATACCGTC	GGCTACCGTT	TCGGCGGCNA	ATATACGGGC	TATTACNGCT
	1451	CGGATGACGA	ATTTAAGCGG	GCATTCCGAG	AAAACCTCGC	GACATACANG
	1501	AAACATTGCA	ACAGAGCTG	CGGAATTTAT	GAACCCGTAT	TGAAAAAATA
	1551	CGGCAAAAAG	CGCGCCAACA	ACCATTCCGT	CAGCATTAGT	GCGGACTTCG
50	1601	GCGATTATTT	CATGCCGTTT	GCCAGCTATT	CGCGCACACA	CCGTATGCCC
	1651	AACATCCAAG	AAATGTATTT	TTCCCAAATC	GGCGACTCCG	GCGTTTCACAC
	1701	CGCCTTAAAA	CCAGAGCGCG	CAAACACTTG	GCAATTTGGC	TTCAATACCT
	1751	ATAAAAAAGG	ATTGTTAAAA	CAAGATGATA	TATTAGGATT	AAAACCTGGT
	1801	GGCTACCGCA	GCCGCATCGA	CNACTACATC	CACAACGTTT	ACGGGAAATG
55	1851	TGGGGATTTG	AACGGGAATA	TTCCGAGCTG	GGTCAGCAGC	ACCGGGCTTG
	1901	CCTACACCAT	CCAACACCGC	AATTTCAAAG	ACAAAGTGCA	CAAACACGGT
	1951	TTTGAGTTGG	AGCTGAATTA	CGATTATNGG	CGTTTTTTTCA	CCAACCTTTC
	2001	TTACGCCTAT	CAAAAAAGCA	CGCAACCGAC	CAACTTCAGC	GATGCGAGCG
	2051	AATCGCCCAA	CAATGCGTCC	AAAGAAGACC	AACTCAAACA	AGGTTATGGG
60	2101	TTGAGCAGGG	TTTCCGCCCT	GCCGCGAGAT	TACGGACGTT	TGGAAGTCGG
	2151	TACGCGCTGG	TGCGGCCAAC	AACCTGACTT	GGGCGGCGCG	ATGCGCTATT
	2201	TCGGCAAGAG	CATCCGCGCG	ACGGCTGAAG	AACGCTATAT	CGACGNCACC
	2251	AATGGGGNAN	NTACCAGCAA	TTTCCGGCAA	CTGGGCAAGC	GTTCCATCAN
	2301	ACAAACCGAA	ACCCTTGCCC	GCCAGCCTTT	GATTTTTTGT	TTNTACGCCG
65	2351	CTTACGAGCC	GAAGAAAAAN	CTTATTTTCC	GCGCCGAAGT	CAAAAATCTG
	2401	TTGCACAGGC	GTTATATCGA	TCCGCTCGAT	GCGGGCAATG	ATGCGGCAAC
	2451	GCAGCGTTAT	TACAGTTCTG	TCGACCCGAA	AGACAAGGAC	GAAGAAGTAA
	2501	CGTGTAATGA	TGATAACACG	TTATGCAACG	GCAAAATACG	CGGCACAAGC
	2551	AAAAGCGTAT	TGACCAATTT	TGCACGCGGA	CNACCTTTT	TGATAACGAT
	2601	GAGCTACAAG	TTTTTAA			

This encodes a protein having (partial) amino acid sequence <SEQ ID 880>:

```

      1  KDKKVFTDAR AVSTRQDIFK SXENLDNIVR XIPGAFTXQX KSSGXVSLNI
    51  RXDSGFGRVN TMVDGITXTF YSTSTDAGRA GGSSQFGASV DSNFXAGLDV
   101  VKGSFSGSAG INSLAGSANL RTLXVDDVVQ GNXTYGLLLK GLTGTNSTKG
   151  NAMAAGARK  WLESGASGV LYGHSRRSVA QNYRVGGGGQ HIGNFGAEYL
   201  ERRKQRYFEQ EGGLKFNSNS GKWERDFQKS YWKTWKYQKY DAPQELQKYI
   251  EGHDKSWREN LAPQYDITPI DPSSLKXQSA GNLFKLEYDG VFNKYTAQFR
   301  DLNTKIGSRK IINRNYQFNY GLSLNPYTNL NLTAAYNSGR QKYPKGSKFT
   351  GWGLXKDFET YNNAKILDIX NTSTFRLPRE TELQTLGFN YFHNEYGKNR
   401  FPEELGLFFD GPDNDGLYS YLGRFKGDKG LLPQKSTIVQ PAGSQYFNTF
   451  YFDAALKKDI YRLNYSTNTV GYRFGGXVTG YYXSDDEFKR AGENSPITYX
   501  KHCNQSCGIY EPVLKKGK  RANNHSVIS ADFGDYFMPF ASYSRTHRMP
   551  NIQEMYFSQI GDSGVHTALK PERANTWQFG FNTYKKGLLK QDDILGLKLV
   601  GYRSRIDXYI HNVYGKWWDL NGNIPSWVSS TGLAYTIQHR NFKDKVHKHG
   651  FELELNIDYX RFFTNLSYAY QKSTQPTNFS DASESPNNAS KEDQLKQGYG
   701  LSRVSALPRD YGRLEVGTWR LGNKLTLLGA MRYFGKSIRA TAEERYIDXT
   751  NGXXTSNFRQ LGKRSIXQTE TLARQPLIFD XYAAYPEKKX LIFRAEVKNL
   801  FDRRIDPLD  AGNDAATQRY YSSFDPKDKD EEVTCNDNT LCNGKYGGTS
   851  KSVLTNFARG XTFLITMSYK F*

```

20 ORF133a and ORF133-1 show 94.3% identity in 871 aa overlap:

```

                                     10      20      30      40
orfl33a.pep                        KDKKVFTDARAVSTRQDIFKSXENLDNIVRXIPGAFTXQXKS
                                     |||
orfl33-1                          EAQIQVLEDVHVKAARVPKDKKVFTDARAVSTRQDIFKSSENLDNIVRSIPGAFTQQDKS
                                     10      20      30      40      50      60

                                     50      60      70      80      90      100
orfl33a.pep                      SGXVSLNIRXDSGFGRVNTMVDGITXTFYSTSTDAGRAGGSSQFGASVDSNFXAGLDVVK
                                     |||
orfl33-1                          SGIVSLNIRGDSGFGRVNTMVDGITQTFYSTSTDAGRAGGSSQFGASVDSNFIAGLDVVK
                                     70      80      90      100      110      120

                                     110     120     130     140     150     160
orfl33a.pep                      GSFGSAGINSLAGSANLRTLXVDDVVQGNXTYGLLLKGLTGTNSTKGNAMAAGARKWL
                                     |||
orfl33-1                          GSFGSAGINSLAGSANLRTLGVDDVVQGNNTYGLLLKGLTGTNSTKGNAMAAGARKWL
                                     130     140     150     160     170     180

                                     170     180     190     200     210     220
orfl33a.pep                      ESGASVGVLYGHSRRSVAQNYRVGGGGQHIGNFGAEYLERRKQRYFEQEGGLKFNSNSGK
                                     |||
orfl33-1                          ESGASVGVLYGHSRRSVAQNYRVGGGGQHIGNFGAEYLERRKQRYFVQEGALKFNSDSGK
                                     190     200     210     220     230     240

                                     230     240     250     260     270     280
orfl33a.pep                      WERDFQKSYWKTWKYQKYDAPQELQKYIEGHDKSWRENLAPOYDITPIDPSSLKXQSAGN
                                     |||
orfl33-1                          WERDLQRQQWKYPYKNYNN-QELQKYIEEHDKSWRENLPQYDITPIDPSSLKXQSAGN
                                     250     260     270     280     290

                                     290     300     310     320     330     340
orfl33a.pep                      LFKLEYDGVFNKYTAQFRDLNTKIGSRKIINRNYQFNYGLSLNPYTNLNLTAAYNSGRQK
                                     |||
orfl33-1                          LFKLEYDGVFNKYTAQFRDLNTKIGSRKIINRNYQFNYGLSLNPYTNLNLTAAYNSGRQK
                                     300     310     320     330     340     350

                                     350     360     370     380     390     400
orfl33a.pep                      YPKGSKFTGWGLXKDFETYNNAKILDIXNTSTFRLPRETELQTLGFNYFHNEYGKNRFP
                                     |||
orfl33-1                          YPKGSKFTGWGLXKDFETYNNAKILDIXNTSTFRLPRETELQTLGFNYFHNEYGKNRFP
                                     360     370     380     390     400     410

                                     410     420     430     440     450     460
orfl33a.pep                      EELGLFFDGPDXDNGLYSYLGRFKGDKGLLPQKSTIVQPAGSQYFNTFYFDAALKKDIYR
                                     |||
orfl33-1                          EELGLFFDGPDXDNGLYSYLGRFKGDKGLLPQKSTIVQPAGSQYFNTFYFDAALKKDIYR
                                     420     430     440     450     460     470

```

-478-

		470	480	490	500	510	520
	orf133a.pep	LNYSTNTVGYRFGGXYTGYYXSDDEFKRAFGENSPTYXKHCNQSCGIYEPVLKKGKRA					
5	orf133-1	LNYSTNTVGYRFGGEYTGYYGSDDEFKRAFGENSPTYXKHCNRSCGIYEPVLKKGKRA	480	490	500	510	520
							530
	orf133a.pep	NNHSVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTALKPERANTWQFGFN					
10	orf133-1	NNHSVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTALKPERANTWQFGFN	540	550	560	570	580
							590
	orf133a.pep	TYKKGLLKQDDILGLKLVGYRSRIDXYIHNVYGKWWDLNGNIPSWVSSTGLAYTIQHRNF					
15	orf133-1	TYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVSSTGLAYTIQHRNF	590	600	610	620	630
							640
			600	610	620	630	640
							650
	orf133a.pep	KDKVHKHGFEELELNLDYXRFFTNLSYAYQKSTQPTNFSDAESPNNASKEDQLKQGYGLS					
20	orf133-1	KDKVHKHGFEELELNLDYGRFFTNLSYAYQKSTQPTNFSDAESPNNASKEDQLKQGYGLS	650	660	670	680	690
							700
			660	670	680	690	700
							710
25	orf133a.pep	RVSALPRDYGRLEVGTWLGKLTGGAMRYFGKSIRATAEERYIDXTNGXXTSNFRQLG					
	orf133-1	RVSALPRDYGRLEVGTWLGKLTGGAMRYFGKSIRATAEERYIDGTNGGNTSNFRQLG	710	720	730	740	750
							760
30			720	730	740	750	760
							770
	orf133a.pep	KRSIXQTETLARQPLIFDXAAYEPKKNLIFRAEVKNLFDRLYIDPLDAGNDAATQRYYS					
35	orf133-1	KRSIXQTETLARQPLIFDFYAAYEPKKNLIFRAEVKNLFDRLYIDPLDAGNDAATQRYYS	770	780	790	800	810
							820
			780	790	800	810	820
							830
	orf133a.pep	SEDPKDKDEEVTNCDDNTLCNGKYGGTSKSVLTNFARGXTFLITMSYKFX					
40	orf133-1	SEDPKDKDEDVTCNADKTLNCGKYGGTSKSVLTNFARGRTFLMTMSYKFX	830	840	850	860	870
							880
			840	850	860	870	880

Homology with a predicted ORF from *N.gonorrhoeae*ORF133 shows 92.3% identity over 392 aa overlap with a predicted ORF (ORF133ng) from *N.*45 *gonorrhoeae*:

	orf133.pep	PGYYGSDDEFKRAFGENSPTYXKHCNRSCGI	31
	orf133ng	FYFDAALKKDIYRLNYSTNAINYRFGGEYTGYYGSENEFKRAFGENSPAYKEHCDCPSGL	560
50	orf133.pep	YEPVLKKGKRRANNHSVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL	91
	orf133ng	YEPVLKKGKRRANNHSVSISADFGDYFMPFAGYSRTHRMPNIQEMYFSQIGDSGVHTAL	620
55	orf133.pep	KPERANTWQFGFXTYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVS	151
	orf133ng	KPERANTWQFGFNTYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVG	680
	orf133.pep	STGLAYTIQHRFXDKVHQXXXXXXXXXYDYGRFFTNLSYAYQKSTQPTNFSDAESPNN	211
60	orf133ng	STGLAYTIRHRNFKDKVHKHGFEELELNLDYGRFFTNLSYAYQKSTQPTNFSDAESPNN	740
	orf133.pep	SKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKLTGGAMRYFGKSIRATAEERYIDG	271
	orf133ng	SKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKLTGGAMRYFGKSIRATAEERYIDG	800
65	orf133.pep	TNGGNTSNFRQLGKRSIXQTETLARQPLIXDFNAAYEPKKNLIFRAEVKNLFDRLYIDPL	331
	orf133ng	TNGGNTSNVRQLGKRSIXQTETLARQPLIFDFYAAYEPKKNLIFRAEVKNLFDRLYIDPL	860

The complete length ORF133ng nucleotide sequence <SEQ ID 881> is predicted to encode a protein having amino acid sequence <SEQ ID 882>:

15

25

35

-480-

1801 CCCAACATCC AAGAAATGTA TTTTCCCAA ATCGGCGACT CCGGCGTTCA  
 1851 CACCGCCTTA AAACCAGAGC GCGCAAACAC TTGGCAATTT GGCTTCAATA  
 1901 CCTATAAAAA AGGATTGTTA AAACAAGATG ATATATTAGG ATTGAAACTG  
 1951 GTCGGCTACC GCAGCCGCAT TGACAACATC ATCCACAACG TTACCGGGAA  
 5 2001 ATGGTGGGAT TTGAACGGGG ATATTCCGAG CTGGGTCGGC AGCACCGGGC  
 2051 TTGCCTACAC CATCCGACAC CGCAATTTCA AAGACAAAGT GCACAAACAC  
 2101 GGTTTTGAGC TGGAGCTGAA TTACGATTAT GGGCGTTTTT TCACCAACCT  
 2151 TTCTTACGCC TATCAAAAAA GCACGCAACC GACCAATTTC AGCGATGCGA  
 10 2201 GCGAATCGCC CAACAATGCC tccaaAGAAG ACCAACTCAA ACAAGGTTAT  
 2251 GGGCTGAGCA GGGTTTCCGC CCTGCCGCGA GATTACGGAC GTTTGGAAGT  
 2301 CGGTACGCGC TGGTTGGGCA ACAAACCTGAC TTTGGGCGGC GCGATgcGCT  
 2351 ATTTCCGCAA GAGCATCCGC GCGACGGCTG AAGAACGCTA TATCGACGGC  
 2401 ACCAACGGGG GAAATACCAG CAATGTCCGG CAACTGGGCA AGCGTTCCAT  
 2451 CAAACAAACC GAAACCCCTG CCCGACAGCC TTTGATTTTT GATTTTTACG  
 15 2501 CCGCTTACGA CCCGAAGAAA AACCTTATTT TCCGCGCGCA AGTCAAAAAC  
 2551 CTGTTGACA GGC GTTATAT CGATCCGCTC GATGCGGGCA ATGATGCGGC  
 2601 AACCGACGCT TATTACAGCT CGTTCGACCC GAAAGACAAG GACGAAGACG  
 2651 TAACGTGTAA TGCTGATAAA AC GTTGTGCA ACGGCAAATA CGGCGGCACA  
 20 2701 AGCAAAAGCG TATTGACCAA TTTGCGACGC GGACGCACCT TCTTGATGAC  
 2751 GATGAGCTAC AAGTTTTAA

This corresponds to the amino acid sequence <SEQ ID 884; ORF133ng-1>:

1 MRSSFRLKPI CFYLMGVMLY HHSYAEDAGR AGSEAQIQVL EDVHVKAHRV  
 51 PKDKKVETDA RAVSTRQDVF KSGENLDNIV RSIPGAFTQQ DKSSGIVSLN  
 101 IRGDSGFGRV NTMVDGITQT FYSTSTDAGR AGGSSQFGAS VDSNFIAGLD  
 151 VVKGSFSGSA GINSLAGSAN LRTLGVDDV QGNNTYGLLL KGLTGTNSTK  
 201 GNMAAIGAR KWLESGASVG VLYGHSRRGV AQNYRVGGGG QHIGNFGEEY  
 251 LERRKQQYFV QEGGLKFNAG SGKWERDLQR QYWKTKWYKK YEDPQELQKY  
 301 IEEHDKSWRE NLAPQYDITP IDPSGLKQOS AGNLFKLEYD GVFNKYTAQF  
 351 RDLNTRIGSR KIINRNYQFN YGLSLNPTYN LNLTAAYNSG RQKYPKGAKF  
 401 TWGGLLKDFE TYNNAKILD L NNTATFRLPR ETELQTTLGF NYFHNEYGKN  
 451 RFPEELGLFF DGPDQDNGLY SYLGRFKGDK GLLPQKSTIV QPAGSQYFNT  
 501 FYFDAALKKD IYRLNYSTNA INYRFGGEYT GYGSSENEFK RAFGENSPAY  
 551 KEHCDPSCGL YEPVLKKYK KRAANHVSUI SADFGDYFMP FAGYSRTHRM  
 601 PNIQEMYFSQ IGDSGVHTAL KPERANTWQF GFNTYKKGLL KQDDILGLKL  
 35 651 VGYRSRIDNY IHNVYKWWD LNGDIPSWVG STGLAYTIRH RNFKDKVHKH  
 701 GFELELNIDY GREFTNLSYA YQKSTQPTNF SDASESPNNA SKEDQLKQGY  
 751 GLSRVSALPR DYGRLEVGT WLGNKLTGG AMRYFGKSIR ATAERYIDG  
 801 TNGGNTSNVR QLGRSRIQT ETLARQPLIF DFYAAYEPK NLI FRAEVKN  
 851 LFDRRYIDPL DAGNDAATQR YYSSFDPKDK DEDVTCNADK T LCNKGKGGT  
 40 901 SKSVLTNFA R GRTFLMTMSY KF\*

ORF133ng-1 and ORF133-1 show 96.2% identity in 889 aa overlap:

10 20 30 40 50 60  
 orf133ng-1.pep SFRLKPICFYLMGVMLYHHSYAEDAGRAGSEAQIQVLEDVHVKAHRVPKDKKVETDARAV  
 45 orf133-1 EAQIQVLEDVHVKAHRVPKDKKVETDARAV  
 10 20 30  
 70 80 90 100 110 120  
 orf133ng-1.pep STRQDVFKSGENLDNIVRSIPGAFTQQDKSSGIVSLNIRGDSGFGRVNTMVDGITQTFYS  
 50 orf133-1 STRQDIFKSSSENLDNIVRSIPGAFTQQDKSSGIVSLNIRGDSGFGRVNTMVDGITQTFYS  
 40 50 60 70 80 90  
 130 140 150 160 170 180  
 orf133ng-1.pep TSTDAGRAGGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDDVQGN  
 55 orf133-1 TSTDAGRAGGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDDVQGN  
 100 110 120 130 140 150  
 190 200 210 220 230 240  
 orf133ng-1.pep NTYGLLLKGLTGTNSTKGNMAAIGARKWLESGASVG VLYGHSRRGVAQNYRVGGGGQHI  
 60 orf133-1 NTYGLLLKGLTGTNSTKGNMAAIGARKWLESGASVG VLYGHSRRGVAQNYRVGGGGQHI  
 160 170 180 190 200 210  
 250 260 270 280 290 300  
 orf133ng-1.pep GNFGEEYLERRKQQYFVQEGGLKFNAGSGKWERDLQRQYWKTKWYKKYEDPQELQKYIEE

70 In addition, ORF133ng-1 is homologous to a TonB-dependent receptor in *H.influenzae*:

-482-

sp|P45114|YC17\_HAEIN PROBABLE TONB-DEPENDENT RECEPTOR H11217 PRECURSOR  
>gi|1075372|pir||G64110 transferrin binding protein 1 precursor (tbpl) homolog -  
Haemophilus influenzae (strain Rd KW20) >gi|1574147 (U32801) transferrin binding  
protein 1 precursor (tbpl) [Haemophilus influenzae] Length = 913

Score = 930 bits (2377), Expect = 0.0

Identities = 476/921 (51%), Positives = 619/921 (66%), Gaps = 72/921 (7%)

Query: 38 QVLEDVHVKAARVPKDKKVFTDARAVSTRQDVFKSGENLDNIVRSIPGAFTQQDKSSGIV 97

+ L + V K + DKK FT+A+A STR++VFK + +D ++RSIPGAFTQQDK SG+V

Sbjct: 29 ETLGQIDVVEKVISNDKKPFTEAKAKSTRENVFKETQTIDQVIRSIPGAFTQQDKSGGV 88

Query: 98 SLNIRGDSGFGRVNTMVDGITQTFYSTSTDAGRAGGSSQFGASVDSNFIAGLDVVKGSFS 157

S+NIRG++G GRVNTMVDG+TQTFYST+ D+G++GGSSQFGA++D NFIAG+DV K +FS

Sbjct: 89 SVNIRGENGLGRVNTMVDGVTQTFYSTALDSGQSGGSSQFGAAIDPNFIAGVDVNKSNFS 148

Query: 158 GSAGINSLAGSANLRTLGVDDVVOXXXXXXXXXXXXXXXXXXXXXAMAAIGARKWLESGA 217

G++GIN+LAGSAN RTLGV+DV+ M RKWL++G

Sbjct: 149 GASGINALAGSANFRTLGVNDVITDDKPFGIILKGMTGSNATKSNFMTMAAGRWLDNGG 208

Query: 218 SVGVLYGHSRRGVAQNYRVGGGQOHIGNFGEYLERRKQYFVQEGGLKFNAGSGKWERD 277

VG+Y+G+S+R V+Q+YR+ GGG+ + + G++ L + K+ YF + G N G+W D

Sbjct: 209 YGVVYGYSGREVSQDYRI-GGGERLASLGQDILAKEKEAYF-RNAGYILNP-EGQWTPD 265

Query: 278 LQRQYWK-----TKWY-----KKYEDPQELQK---YIEE 303

L +++W +Y KK +D ++LQK IEE

Sbjct: 266 LSKKHWSCNKPDPYQKNGDCSYRIGSAAKTRREILQELLTNGKKPKDIEKLQKNGDGIEE 325

Query: 304 HDKSWRENLAPOYDITPIDPSGLKQOSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII 363

DKS+ N QY + PI+P L+ +S +L K EY AQ R L+ +IGSRKI

Sbjct: 326 TDKSFERN-KDQYSVAPIEPGSLQSRSRSHLLKFEYGDHQNGLAQLRTLNDKIGSRKIE 384

Query: 364 NRNYQFNYSLSNPYTNLNLTAAYNSGRQKYPKGAFTGWGLLKDFETYNNAKILDNLNT 423

NRNYQ NY + N Y +LNL AA+N G+ YPKG F GW + T N A I+D+NN+

Sbjct: 385 NRNYQVNYNFNNNSYLDNLMAAHNIGKTIYPKGGFFAGWQVADKLITKNVANIVDINNS 444

Query: 424 ATFRLPRETELQTTLGFNYPHNEYGKNRFPEELGLFFDGPDDQDNGLYSY--LGRFGKDGK 481

TF LP+E +L+TTLGFNYP NEY KNRFPPEEL LF++ D GLYS+ GR+ G K

Sbjct: 445 HTFLLPKEIDLKTTLGFNYPFTNEYSKNRFPEELSLFYNDASHDQGLYSHSKRGYSGTGS 504

Query: 482 LLPQKSTIVQFAGSQYFNTFYFDAALKKDIYRLNYSNAINYRFGGEYTGYYGSENEFKR 541

LLPQ+S I+QP+G Q F T YFD AL K IY LNYS N +Y F GEY GY

Sbjct: 505 LLPQRSVILQPSGKQKFKTVYFDALTALSKGIYHLNYSVNFTHYAFNGEYVGY----- 555

Query: 542 AFGENSPAYKEHCDPSCGLYE PVLKKGKRRANNHVSISADFGDYFMPFAGYSRTHRMP 601

EN+ + + EP+L K G K+A NHS ++SA+ DYFMPF YSRTHRMP

Sbjct: 556 ---ENTAGQQ-----INEPILHKSGHKAFNHSATLSAELSDYFMPFFTYSRTHRMP 604

Query: 602 NIEQMYFSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDILGLKLVGYRSRIDNYI 661

NIEQ+FSQ+ ++GV+TALKPE++T+Q GFNTYKKG L QDD+LG+KLVGYRS I NYI

Sbjct: 605 NIEQMFSSQVSNAGVNTALKPEQSDTYQLGFNTYKKG LFTQDDVLGVKLVGYRSFIKNI 664

Query: 662 HNVYKWWDLNGDIPSWVGSTGLAYTIRHRNFKDKVHKHGFELNLDYGRFFTNLSYAY 721

HNVYG WW +P+W S G YTI H+N+K V K G ELE+NYD GRFF N+SYAY

Sbjct: 665 HNVYGVWW--RDGMPTWAESNGFKYTIHQNYKPIVKKSGVELEINYDMGRFFANVSAY 722

Query: 722 QKSTQPTNFSDAESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKLTGGA 781

Q++ QPTN++DAS PNNAS+ED LKQGYGLSRVS LP+DYGRLE+GTRW KLTG A

Sbjct: 723 QRTNQPTNYADASPRPNNASQEDILKQGYGLSRVSMLPKDYGRLELGTRWFDQKLTGLA 782

Query: 782 MRYFGKSIRATAERYIDGTNGGNTSNVRQLGKRSIKQTETLARQPLIFDYAAYEPKKN 841

RY+GKS RAT EE YI+G+ + +R+ ++K+TE + +QP+I D + +YEP K+

Sbjct: 783 ARYYGKSKRATIEEYINGSR-FKKNLRRRENYAVKKTEDIKKQPIILDHVSYPEIKD 841

Query: 842 LIFRAEVKNLFDRIYIDPLDAGNDAATQRYSSFPDKDKDEDVTCNADKTLNKGKGGTS 901

LI +AEV+NL D+RY+DPLDAGNDA+QRYSS + + C D + C GG+

Sbjct: 842 LIIKAEVQNLLDKRYVDPLDAGNDAASQRYSSSL-----NNSIECAQDSSAC-----GGSD 892

Query: 902 KSVLTNFARGRTFLMTMSYKF 922

K+VL NFARGRT+++++YKF

Sbjct: 893 KTVLYNFARGRTYILSLNYKF 913

The underlined motif in the gonococcal protein (also present in the meningococcal protein) is predicted to be an ATP/GTP-binding site motif A (P-loop), and the analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

## 5 Example 104

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 885>

```

1  ATGAACCTGA TTTCACGTTA CATCATCCGT CAAATGGCGG TTATGGCGGT
51  TTACGCGCTC CTTGCCTTCC TCGCTTTGTA CAGCTTTTTT GAAATCCTGT
101 ACGAAACCGG CAACCTCGGC AAAGGCAGTT ACGGCATATG GGAAATGCTG
151 GGTACACCG CCCTCAAAT GCCCGCCGCG GCCTACGAAC TGATTCCCCT
201 CGCCGTCCTT ATCGGCGGAC TGGTCTCCCT CAGCCAGCTT GCCGCCGCA
251 GCGAACTGAC CGTCATCAA GCCAGCGGCA TGAGCACCAA AAAGCTGCTG
301 TTGATTCTGT CGCAGTTCGG TTTTATTTT GCTATTGCCA CCGTCGCGCT
351 CGGCGAATGG GTTGCGCCCA CACTGAGCCA AAAAGCCGAA AACATCAAAG
15  401 CCGCCGCCAT CAACGGCAA ATCAGCACCG GCAATACCGG CCTTGGGCTG
451 AAAGAAAAAA ACAGCGTGAT CAATGTGCGC GAAATGTTGC CCGACCAT..

```

This corresponds to the amino acid sequence <SEQ ID 886; ORF112>:

```

1  MNLISRYIIR QMAVMAVYAL LAFLALYSFF EILYETGNLG KGSYGIWEML
51  GYTALKMPAR AYELIPLAVL IGGVLVLSQL AAGSELTVIK ASGMSTKKLL
20  101 LILSQFGFIF AIATVALGEW VAPTLSQKAE NIKAAAIN GK ISTGNTGLWL
151  KEKNSVINVR EMLPDH...

```

Further work revealed further partial nucleotide sequence <SEQ ID 887>:

```

1  ATGAACCTGA TTTCACGTTA CATCATCCGT CAAATGGCGG TTATGGCGGT
51  TTACGCGCTC CTTGCCTTCC TCGCTTTGTA CAGCTTTTTT GAAATCCTGT
25  101 ACGAAACCGG CAACCTCGGC AAAGGCAGTT ACGGCATATG GGAAATGCTG
151  gGTACACCG CCCTCAAAT GCCCGCCGCG GCCTACGAAC TGATTCCCCT
201  CGCCGTCCTT ATCGGCGGAC TGGTCTCCCT CAGCCAGCTT GCCGCCGCA
251  GCGAACTGAC CGTCATCAA GCCAGCGGCA TGAGCACCAA AAAGCTGCTG
301  TTGATTCTGT CGCAGTTCGG TTTTATTTT GCTATTGCCA CCGTCGCGCT
30  351 CGGCGAATGG GTTGCGCCCA CACTGAGCCA AAAAGCCGAA AACATCAAAG
401  CCGCCGCCAT CAACGGCAA ATCAGCACCG GCAATACCGG CCTTGGGCTG
451  AAAGAAAAAA ACAGC rTKAT CAATGTGCGC GAAATGTTGC CCGACCATAC
501  GCTTTTGGGC ATCAAATTT GGGCGCGCAA CGATAAAAC GAATTGCGAG
551  AGGCAGTGGA AGCCGATTCC GCCGTTTGA ACAGCGACGG CAGTTGGCAG
35  601  TTGAAAAACA TCCGCCGCAG CACGCTTGGC GAAGACAAAG TCGAGGTCTC
651  TATTGCGGCT GAAGAAACT GGCCGATTTC CGTCAAACGC AACCTGATGG
701  ACGTATTGCT CGTCAAACCC GACCAAATGT CCGTCGGCGA ACTGACCACC
751  TACATCCGCC ACCTCCAAA CAACAGCCAA AACACCCGAA TCTACGCCAT
801  CGCATGGTGG CGCAAATTGG TTTACCCGCG CGCAGCCTGG GTGATGGCGC
40  851  TCGTCGCCTT TGCCTTTACC CCGCAAACCA CCCGCCACGG CAATATGGGC
901  TTAATACTCT TCGGCGGCAT CTGTstCGGA TTGCTGTTCC ACCTTGCCGG
951  ACGGCTCTTT GGGTTTACCA GCCAACTCG...

```

This corresponds to the amino acid sequence <SEQ ID 888; ORF112-1>:

```

1  MNLISRYIIR QMAVMAVYAL LAFLALYSFF EILYETGNLG KGSYGIWEML
45  51  GYTALKMPAR AYELIPLAVL IGGVLVLSQL AAGSELTVIK ASGMSTKKLL
101  LILSQFGFIF AIATVALGEW VAPTLSQKAE NIKAAAIN GK ISTGNTGLWL
151  KEKNSXINVR EMLPDHTLLG IKIWARNDKN ELAEAVEADS AVLNSDGSWQ
201  LKNIRSTLG EDKVEVSIAA EENWPISVKR NLMDVLLVKP DQMSVGELTT
251  YIRHLQNSQ NTRIYIAIWW RKLVPAAAW VMALVAFAT PQTTRHGNMG
50  301  LKLFGGICXG LLEHLAAGRLF GFTSQL...

```

Computer analysis of this amino acid sequence predicts two transmembrane domains and gave the following results:



Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF112 shows 96.4% identity over a 166aa overlap with an ORF (ORF112a) from strain A of *N.*

*meningitidis*:

5	orf112.pep	10	20	30	40	50	60
	orf112a	10	20	30	40	50	60
10	orf112.pep	70	80	90	100	110	120
	orf112a	70	80	90	100	110	120
15	orf112.pep	130	140	150	160		
	orf112a	130	140	150	160	170	180
20	orf112a	190	200	210	220	230	240

The ORF112a nucleotide sequence <SEQ ID 889> is:

25	1	ATGAACCTGA	TTTCACGTTA	CATCATCCGT	CAAATGGCGG	TTATGGCGGT
	51	TTACGCGCTC	CTTGCCCTTC	TCGCTTTGTA	CAGCTTTTTT	GAAATCCTGT
30	101	ACGAAACCGG	CAACCTCGGC	AAAGGCAGTT	ACGGCATATG	GGAAATGNTG
	151	GGNTACACCG	CCCTCAAAAT	GNCCGCCCGC	GCCTACGAAC	TGATGCCCTT
35	201	CGCCGTCCTT	ATCGGCGGAC	TGGTCTCTNT	CAGCCAGCTT	GCCGCCGGCA
	251	GCGAACTGAN	CGTCATCAAA	GCCAGCGGCA	TGAGCACCAA	AAAGCTGCTG
40	301	TTGATTCTGT	CGCAGTTCGG	TTTTATTTTT	GCTATTGCCA	CCGTCGCGCT
	351	CGGCGAATGG	GTTGCGCCCA	CACTGAGCCA	AAAAGCCGAA	AACATCAAAG
45	401	CCGCGGCCAT	CAACGGCAAA	ATCAGTACCG	GCAATACCGG	CCTTTGGCTG
	451	AAAGAAAAAA	ACAGCATTAT	CAATGTGCGC	GAAATGTTGC	CCGACCATAC
50	501	CCTGCTGGGC	ATTAAATCT	GGGCCCGCAA	CGATAAAAAC	GAAGTGGCAG
	551	AGGCAGTGGA	AGCCGATTCC	GCCGTTTTGA	ACAGCGACGG	CAGTTGGCAG
55	601	TTGAAAAACA	TCCGCCCGAG	CACGCTTGGC	GAAGACAAAG	TCGAGGTCTC
	651	TATTGCGGCT	GAAGAAAANT	GGCCGATTTC	CGTCAAACGC	AACCTGATGG
60	701	ACGTATTGCT	CGTCAAACCC	GACCAAATGT	CCGTCGGCGA	ACTGACCACC
	751	TACATCCGCC	ACCTCCAAAN	NNACAGCCAA	AACACCCGAA	TCTACGCCAT
65	801	CGCATGGTGG	CGCAAATTGG	TTTACCCCGC	CGCAGCCTGG	GTGATGGCGC
	851	TCGTCGCCTT	TGCCTTTACC	CCGCAAACCA	CCCGCCACGG	CAATATGGGC
70	901	TTAAANTCT	TCGGCGGCAT	CTGTCTCGGA	TTGCTGTTCC	ACCTTGCCGG
	951	NCGGCTCTTC	NGGTTTACCA	GCCAACTCTA	CGGCATCCCG	CCCTTCCTCG
75	1001	NCGGCGCACT	ACCTACCATA	GCCTTCGCCT	TGCTCGCCGT	TTGGCTGATA
	1051	CGCAAACAGG	AAAAACGCTA	A		

This encodes a protein having the amino acid sequence <SEQ ID 890>:

50	1	MNLISRYIIR	QMAVMVYAL	LAFLALYSFF	EILYETGNLG	KGSYGIWEMX
	51	GYTALKMXAR	AYELMPLAVL	IGGLVSXSQ	AAGSELXVIK	ASGMSTKKLL
55	101	LILSQFGFIF	AIATVALGEW	VAPTLSQKAE	NIKAAAINGK	ISTGNTGLWL
	151	KEKNSIINVR	EMLPDHTLLG	IKIWARNDKN	ELAEAVEADS	AVLNSDGSWQ
60	201	LKNIRIRSTLG	EDKVEVSIAA	EEXWPISVKR	NLMDVLLVKP	DQMSVGELTT
	251	YIRHLQXXSQ	NTRIYAIAWW	RKLVYPAAAW	VMALVAFAPT	PQTTRHGNMG
65	301	LKXFGGICLG	LLFHLAAGRLF	XFTSQLYGIP	PFLXGALPTI	AFALLAVWLI
	351	RKQEK*				

ORF112a and ORF112-1 show 96.3% identity in 326 aa overlap:

60	orf112a.pep	MNLISRYIIRQMAVMVYALLAFLALYSFFEILYETGNLGKGSYGIWEMXGYTALKMXAR
	orf112-1	MNLISRYIIRQMAVMVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR
	orf112a.pep	AYELMPLAVLIGGLVSXSQLAAGSELXVIKASGMSTKKLLILSQFGFIFAIATVALGEW

	orf112-1	AYELIPLAVLIGGLVLSLSQLAAGSELTVIKASGMSTKKLLLLLSQFGFI FAIATVALGEW
5	orf112a.pep	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSI INVREMLPDHTLLGIKIWARNDKN
	orf112-1	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSXINVREMLPDHTLLGIKIWARNDKN
	orf112a.pep	ELAEAVEADSAVLNSDGSWQLKNIRRSTLGEDKVEVSIAAEEXWPI SVKRNLMDEVLLVKP
10	orf112-1	ELAEAVEADSAVLNSDGSWQLKNIRRSTLGEDKVEVSIAAENWPI SVKRNLMDEVLLVKP
	orf112a.pep	DOMSVGELTTYIRHLQXXSQNTRIYAI AWWRKLVYPAAAWVMALVAFAPTQOTTRHGNMG
	orf112-1	DOMSVGELTTYIRHLQNNSQNTRIYAI AWWRKLVYPAAAWVMALVAFAPTQOTTRHGNMG
15	orf112a.pep	LKXFGGICLGLLFHLAGRLFXFTS QLYGIPFLXGALPTIAFALLAVWLIRKQEKRX
	orf112-1	LKLFGGICXGLLFHLAGRLFGFTSQL

20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF112 shows 95.8% identity over 166aa overlap with a predicted ORF (ORF112ng) from *N.*

**gonorrhoeae:**

	orf112.pep	MNLISRYIIIRQMAVMVAYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR	60
25	orf112ng	MNLISRYIIIRQMAVMVAYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR	60
	orf112.pep	AYELIPLAVLIGGLVLSQLAAGSELTVIKASGMSTKKLLLILSQFGFIFAIATVALGEW	120
30	orf112ng	AYELMPLAVLIGGLASLSQLAAGSELAVIKASGMSTKKLLLILSQFGFIFAAVALGEW	120
	orf112.pep	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSVINVREMLPDH	166
	orf112ng	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKTSIINVRGMLPDHTLLGIKIWARNDKN	180

The complete length ORF112ng nucleotide sequence <SEQ ID 891> is:

35	1	ATGAACCTGA	TTTCACGTTA	CATCATCCGC	CAAATGGCGG	TTATGGCGGT
	51	TTACGCGCTC	CTTGCGTTCC	TCGCTTTGTA	CAGCTTTTTT	GAAATCCTGT
	101	ACGAAACCGG	CAACCTCGGC	AAAGGCAGTT	ACGGCATATG	GGAAATGCCTG
	151	GGCTACACCG	CCCTCAAAAT	GCCCCGCCGC	GCCTACGAAC	TCATGCCCTT
	201	CGCCGTCCTC	ATCGGCGGAC	TGGCCTCTCT	CAGCCAGCTT	GCCGCCGGCA
40	251	GCGAACTGGC	CGTCATCAAA	GCCAGCGGCA	TGAGCACCAA	AAAGCTGCTG
	301	TTGATTCTGT	CTCAGTTCGG	TTTTATTTTT	GCTATTGCCG	CCGTCGCGCT
	351	CGGCGAATGG	GTTGCGCCCA	CGCTGAGCCA	AAAAGCCGAA	AACATCAAag
	401	cCGCGCCCAT	taacggCAAA	ATCAGCAccg	gcAATACCGG	CCTTTggcTG
	451	AAAGAAAAAA	ccAGCATTAT	CAATGTGcGc	GGAATGTTGC	CCGACCATAC
45	501	GCTTTTGGGC	ATCAAAATTT	GGGCGCGCAA	CGATAAAAAAC	GAATTGGCAG
	551	AGGCAGTGGa	AGCCGATTCC	GCCGTTTTGA	ACAGCGACGG	CAGCTGGCAG
	601	TTGAAAAACA	TCCGCCGCAG	CATCATGGGT	ACAGACAAAA	TCGAAACATC
	651	cgCCGCCGCC	GAAGAAACTT	gCGCGATTGC	CGTCAGACGC	AACCTGATGG
	701	ACGTATTGCT	CGCTCAAGCC	GACCAAATGT	CGCTCGGCGA	GCTGACCACC
50	751	TACATCCGCC	ACCTCCAAAA	CAACAGCCAA	AACACCCAAA	TCTACGCCAT
	801	CGCATGGTGG	CGTAAACTCG	TTTACCCCGT	CGCCGCATGG	GTCATGGCGC
	851	TCGTTGCCCT	CGCCTTTACG	CCGCAAACCA	CGCGCCACGG	CAATATGGGG
	901	TTAAAACTCT	TCGGCGGCAT	CTGTCTCGGA	TTGCTGTTCC	ACCTTGCCGG
	951	CAGGCTCTTC	GGGTTTACCA	GCCAACTCTA	CGGCACCCCA	CCCTTCTCTG
55	1001	CCGGCTGCAT	GCTTACCATA	GCCTTCGCCT	TGCTCGCTGT	TTGGCTGATA
	1051	CGCAAACAGG	AAAAACGTTG	A		

This encodes a protein having amino acid sequence <SEQ ID 892>:

	1	<u>MNLISRYIIR</u>	<u>QMAVMVAYAL</u>	<u>LAFLALYSFF</u>	<u>EILYETGNLG</u>	<u>KGSYGIWEML</u>
	51	<u>GYTALKMPAR</u>	<u>AYELMPLAVL</u>	<u>IGGLASLSQL</u>	<u>AAGSELAVIK</u>	<u>ASGMSTKKLL</u>
60	101	<u>LILSQFGFIF</u>	<u>AIAAVALGEW</u>	<u>VAPTLSQKAE</u>	<u>NIKAAAIN GK</u>	<u>ISTGNTGLWL</u>
	151	<u>KEKTSIINVR</u>	<u>GMLPDHTLLG</u>	<u>IKIWARNDKN</u>	<u>ELAEAVEADS</u>	<u>AVLNSDGSWG</u>
	201	<u>LKNIRRSIMG</u>	<u>TDKIETSAAA</u>	<u>EETWPIAVRR</u>	<u>NLMDVLLVKP</u>	<u>DQMSVGELTT</u>
	251	<u>YIRHLQNNSQ</u>	<u>NTQIYAIAWW</u>	<u>RKLVPVPAAW</u>	<u>VMALVAFAPT</u>	<u>PQTRRHGNMG</u>
	301	<u>LKLEGGICLG</u>	<u>LLFHLAGRLF</u>	<u>GFTSQLYGTF</u>	<u>PFLAGALPTI</u>	<u>AFALLAVWLI</u>

351 RKQEKR\*

ORF112ng and ORF112-1 show 94.2% identity in 326 aa overlap:

		10	20	30	40	50	60
5	orf112ng	MNLISRYIIRQMAVM	AVYALLAFLALYSF	FEILYETGNLGKGS	YGIWEMLGYTALKM	PAR	
	orf112-1	MNLISRYIIRQMAVM	AVYALLAFLALYSF	FEILYETGNLGKGS	YGIWEMLGYTALKM	PAR	
		10	20	30	40	50	60
10	orf112ng	AYELMPLAVLIGGL	ASLSQLAAGSELAV	IKASGMSTKKLL	LILSQFGFIFAIAA	VALGEW	
	orf112-1	AYELIPLAVLIGGL	VLSLSQLAAGSELT	VIKASGMSTKKLL	LILSQFGFIFAIA	TVALGEW	
		70	80	90	100	110	120
15	orf112ng	VAPTLSQKAENIKAA	AINKGISTGNTGL	WLKEKTSIINVRG	MLPDHTLLGIKI	WARNDKN	
	orf112-1	VAPTLSQKAENIKAA	AINKGISTGNTGL	WLKEKNSXINVRE	MLPDHTLLGIKI	WARNDKN	
		130	140	150	160	170	180
20	orf112ng	ELAEAVEADSAVLN	SDGSWQLKNIRRS	IMGTDKIETSAAA	EETWPIAVRRNL	MDVLLVKP	
	orf112-1	ELAEAVEADSAVLN	SDGSWQLKNIRRS	TLGEDKVEVSIAA	EENWPISVKNR	NLMDVLLVKP	
		190	200	210	220	230	240
25	orf112ng	DQMSVGELTTYIRH	LQNNNQNTQIYAI	AWWRKLVYPVAA	WVMAVAFATPQT	TRHGNMG	
	orf112-1	DQMSVGELTTYIRH	LQNNNQNTQIYAI	AWWRKLVYPAA	WVMAVAFATPQT	TRHGNMG	
		250	260	270	280	290	300
30	orf112ng	LKLFGGICLGLLFH	LAGRLFGFTSQL	YGTTPFLAGALPT	IAFALLAVWLIR	KQEKRX	
	orf112-1	LKLFGGICXGLLFH	LAGRLFGFTSQL				
		310	320	330	340	350	

This analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their  
 40 epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

It will be appreciated that the invention has been described by means of example only, and that  
 modifications may be made whilst remaining within the spirit and scope of the invention.

TABLE I – PCR primers

ORF	Primer	Sequence	Restriction sites
ORF 1	Forward	CGCGGATCCGCTAGC-GGACACACTTATTTTCGG	BamHI-NheI XhoI
	Reverse	CCCGCTCGAG-CCAGCGGTAGCCTAATT	
ORF 2	Forward	GCGGATCCCATATG-TTTGATTTTCGGTTTGGG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GACGGCATAACGGCG	
ORF 2-1	Forward	GCGGATCCCATATG-TTTGATTTTCGGTTTGGG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TGATTTACGGACGCGCA	
ORF 4	Forward	GCGGATCCCATATG-TGCGGAGGTCAAAAAGAC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TTTGGCTGCGCCTTC	
ORF 5	Forward	GGAATTCATATGGCCATGG-TGGAAGGCGCACAACC	NdeI-NcoI BamHI XhoI
	Reverse	CGGGATCC-ATGGAAGGCGCACAAC CCCGCTCGAG-GACTGTGCAAAAACGG	
ORF 6	Forward	CGCGGATCCCATATG-ACCCGTCAATCTCTGCA	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TGCGCCGAACACTTTC	
ORF 7	Forward	CGCGGATCCGCTAGC-GCGCTGCTTTTTGTTC	BamHI-NheI XhoI
	Reverse	CCCGCTCGAG-TTTCAAAATATATTTGCGGA	
ORF 8	Forward	GCGGATCCCATATG-GCTCAACTGCTTCGTAC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-AGCAGGCTTTGGCGC	
ORF 9	Forward	CGCGGATCCCATATG-CCGAAGGAAGTCGGAAA	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TTTCGAGTTTTCGGG	
ORF 10	Forward	GCGGATCCCATATG-GACACAAAAGAAATCCTC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TAATGGGAACCTTGTTTT	
ORF 11	Forward	GCGGATCCCATATG-GCGGTCAACCTCTACG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GGAAACGACTTCGCC	
ORF 13	Forward	CGCGGATCCCATATG-GCTCTGCTTTCCGCGC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-AGGGTGTGTGATAATAAG	
ORF 15	Forward	GGAATTCATATGGCCATGG-GCGGGACACTGACAG	NdeI-NcoI BamHI XhoI
	Reverse	CGGGATCC-TGCGGGACACTGACAGG CCCGCTCGAG-AGGTTGGCCTTGCTATG	
ORF 17	Forward	GGAATTCATATGGCCATGG -TTGCCGGCCTGTTCG	NdeI-NcoI

	Forward	CGGGATCC-ATTGCCGGCCTGTTCG	BamHI
	Reverse	CCCGCTCGAG-AAGCAGGTTGTACAGC	XhoI
<b>ORF 18</b>	Forward	GCGGATCCCATATG-ATTTTGTCTGCATTGGAT	BamHI-NdeI
	Reverse	CCCGCTCGAG-TCTTCCAATTTCTGAAAGC	XhoI
<b>ORF 19</b>	Forward	GGAATTCCATATGGCCATGG -TCGCCAGTGT TTTTACC	NdeI-NcoI
	Forward	CGGGATCC-TTCGCCAGTGT TTTTACCG	BamHI
	Reverse	CCCGCTCGAG-GGTGT TTTTGAAGCTGCC	XhoI
<b>ORF 20</b>	Forward	GGAATTCCATATGGCCATGG -TCGGCGCGGGTATG	NdeI-NcoI
	Forward	CGGGATCC-TTCGGCGCGGGTATG	BamHI
	Reverse	CCCGCTCGAG-CGGCGAGCGAGAGCA	XhoI
<b>ORF 22</b>	Forward	GGAATTCCATATGGCCATGG-TGATTAAATCAAAAAGGTCT	NdeI-NcoI
	Forward	CGGGATCC-ATGATTAAATCAAAAAGGTCTAAACC	BamHI
	Reverse	CCCGCTCGAG-ATTATGATAGCGGCCC	XhoI
<b>ORF 23</b>	Forward	CGCGGATCCCATATG-GATGTTTCTGTTTCAGAC	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTAAACCGATAGGTAAACG	XhoI
<b>ORF 24</b>	Forward	GGAATTCCATATGGCCATGG -TGATGCCGGAATGGTG	NdeI-NcoI
	Forward	CGGGATCC-ATGATGCCGGAATGGTG	BamHI
	Reverse	CCCGCTCGAG-TGTCAGCGTGCGCA	XhoI
<b>ORF 25</b>	Forward	GCGGATCCCATATG-TATCGCAAATGATTGC	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATCGATGGAATAGCCG	XhoI
<b>ORF 26</b>	Forward	GCGGATCCCATATG -CAGCTGATCGACTATTC	BamHI-NdeI
	Reverse	CCCGCTCGAG-GACATCGGCGGTTTT	XhoI
<b>ORF 27</b>	Forward	GGAATTCCATATGGCCATGG-AGACCTATTCTGTTTA	NdeI-NcoI
	Forward	CGGGATCC- CAGACCTATTCTGTTTATTTTAATC	BamHI
	Reverse	CCCGCTCGAG-GGGTTCGATTAAATAACCAT	XhoI
<b>ORF 28</b>	Forward	GGAATTCCATATGGCCATGG-ACGGCTGTACGTTGATGT	NdeI-NcoI
	Forward	CGGGATCC-AACGGCTGTACGTTGATG	BamHI
	Reverse	CCCGCTCGAG-TTTGT CAGAGGAATTCGCG	XhoI
<b>ORF 29</b>	Forward	GCGGATCCCATATG -AACGGTTTGGATGCCCCG	BamHI-NdeI
	Forward	CGCGGATCCGCTAGC-AACGGTTTGGATGCCCCG	BamHI-NheI
	Reverse	CCCGCTCGAG-TTTGTCTAAGTTCTCTGATATG	XhoI
<b>ORF 32</b>	Forward	CGCGGATCCCATATG-AATACTCCTCCTTTTG	BamHI-NdeI
	Reverse	CCCGCTCGAG-GCGTATTTTGTATGCTTTG	XhoI
<b>ORF 33</b>	Forward	GCGGATCCCATATG -ATTGATAGGGATCGTATG	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTGATCTTCAACGGCC	XhoI

<b>ORF 35</b>	Forward Forward Reverse	GCGGATCCCATATG-TTCAGAGCTCAGCTT CGCGGATCCGCTAGC-TTCAGAGCTCAGCTT CCCGCTCGAG-AAACAGCCATTTGAGCGA	BamHI-NdeI BamHI-NheI XhoI
<b>ORF 37</b>	Forward Reverse	GCGGATCCCATATG-GATGACGTATCGGATTTT CCCGCTCGAG-ATAGCCCGCTTTTCAGG	BamHI-NdeI XhoI
<b>ORF 58</b>	Forward Reverse	CGCGGATCCGCTAGC-TCCGAACGCGAGTGGAT CCCGCTCGAG-AGCATTTGTCCAAGGGGAC	BamHI-NheI XhoI
<b>ORF 65</b>	Forward  Forward Reverse	GGAATTCCATATGGCCATGG -TGCTGTATCTGAATCAAG  CGGGATCC-TTGCTGTATCTGAATCAAGG CCCGCTCGAG-CCGCATCGGCAGACA	NdeI-NcoI BamHI XhoI
<b>ORF 66</b>	Forward Reverse	GCGGATCCCATATG-TACGCATTTACCGCCG CCCGCTCGAG-TGGATTTTGCAGAGATGG	BamHI-NdeI XhoI
<b>ORF 72</b>	Forward Reverse	CGCGGATCCCATATG- AATGCAGTAAAAATATCTGA CCCGCTCGAG-GCCTGAGACCTTTGCAA	BamHI-NdeI XhoI
<b>ORF 73</b>	Forward Reverse	GCGGATCCCATATG-AGATTTTTCGGTATCGG CCCGCTCGAG-TTCATCTTTTTCATGTTTCG	BamHI-NdeI XhoI
<b>ORF 75</b>	Forward Reverse	GCGGATCCCATATG- TCTGTCTTTCAAACGGC CCCGCTCGAG-TTTGTTTTTGAAGACAG	BamHI-NdeI XhoI
<b>ORF 76</b>	Forward Reverse	GATCAGCTAGCCATATG-AAACAGAAAAAACCGC CGGGATCC-TTACGGTTTGACACCGTT	NheI-NdeI BamHI
<b>ORF 79</b>	Forward Reverse	CGCGGATCCCATATG-GTTTCCGCCGCCG CCCGCTCGAG-GTGCTGATGCGCTTCG	BamHI-NdeI XhoI
<b>ORF 83</b>	Forward Reverse	GCGGATCCCATATG-AAAACCTGCTGCTGC CCCGCTCGAG-GCCGCCTTTGCGGC	BamHI-NdeI XhoI
<b>ORF 84</b>	Forward Reverse	GCGGATCCCATATG-GCAGAGATCTGTTTG CCCGCTCGAG-GTTTGCCGATCCGACCA	BamHI-NdeI XhoI
<b>ORF 85</b>	Forward Reverse	CGCGGATCCCATATG- GCGGTTTGGGGCGGA CCCGCTCGAG-TCGGCGCGGCGGGC	BamHI-NdeI XhoI
<b>ORF 89</b>	Forward Forward Reverse	GGAATTCCATATGGCCATGG-CCATACCTTCTTATCA CGGGATCC-GCCATACCTTCTTATCAGAG CCCGCTCGAG-TTTTTTGCATTAGAAAAAGC	NdeI-NcoI BamHI XhoI
<b>ORF 97</b>	Forward	GCGGATCCCATATG-CATCCTGCCAGCGAAC	BamHI-NdeI

	Reverse	CCCGCTCGAG-TTCGCCTACGGTTTTTTG	XhoI
ORF 98	Forward	GCGGATCCCATATG-ACGGTAACTGCGG	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTGTTGTTGCGGGCAAATC	XhoI
ORF 100	Forward	GCGGATCCCATATG-TCGGGCATTTACACCG	BamHI-NdeI
	Reverse	CCCGCTCGAG-ACGGGTTTCGGCGGAA	XhoI
ORF 101	Forward	GCGGATCCCATATG-ATTTATCAAAGAAACCTC	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTTCCGCCTTTCAATGT	XhoI
ORF 102	Forward	GCGGATCCCATATG-GCAGGGCTGTTTTACC	BamHI-NdeI
	Reverse	CCCGCTCGAG-AAACGGTTTGAACACGAC	XhoI
ORF 103	Forward	GCGGATCCCATATG-AACCACGACATCAC	BamHI-NdeI
	Reverse	CCCGCTCGAG-CAGCCACAGGACGGC	XhoI
ORF 104	Forward	GCGGATCCCATATG-ACGTGGGGAACGC	BamHI-NdeI
	Reverse	CCCGCTCGAG-GCGGCGTTTGAACGGC	XhoI
ORF 105	Forward	GCGGATCCCATATG-ACCAAATTTCAAACCCCTC	BamHI-NdeI
	Reverse	CCCGCTCGAG-TAAACGAATGCCGTCCAG	XhoI
ORF 106	Forward	GCGGATCCCATATG-AGGATAACCGACGGCG	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTGTTCCCGATGATGT	XhoI
ORF 109	Forward	GCGGATCCCATATG-GAAGATTTATATATAATACTCG	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATCAGCTTCGAACCGAAG	XhoI
ORF110	Forward	AAAGAATTC-ATGAGTAAATCCCGTAGATCTCCC	EcoRI
	Reverse	AAACTGCAG-GGAAAACCATCCGCACTCTGCC	PstI
ORF111	Forward	AAAGAATTC-GCACCACAAAAGGCAAAAACCGCA	EcoRI
	Reverse	AAACTGCAG-TCTGCGCGT TTTGCGGCAGGGTGG	PstI
ORF113	Forward	AAAGAATTC-ATGAACAAAACCTCTATCGTGTGATTTTCAACCG	EcoRI
	Reverse	AAACTGCAG-TTACGAATGCCTGCTTGCTCGACCGTACTG	PstI
ORF115	Forward	AAAGAATTC-TTGCTTGTGCAAACAGAAAAAGACGG	EcoRI
	Reverse	AAAAAGTCGAC-CTATTTT TAGGGGCTTTGCTTGTGTTGAAAAGCCTGCC	Sall
ORF119	Forward	AAAGAATTC-TACAACATGTATCAGGAAACCAATACCG	EcoRI
	Reverse	AAACTGCAG-TTATGAAAACAGGCGCAGGGCGGTTTTGCC	PstI
ORF120	Forward	AAAGAATTC-GCAAGGCTACCCCAATCCGCCGTG	EcoRI
	Reverse	AAACTGCAG-CGGTTTGGCTGCCTGGCCGTTGAT	PstI
ORF121	Forward	AAAGAATTC-GCCTTGGTCTGGCTGGTTTTCGC	EcoRI
	Reverse	AAACTGCAG-TCATCCGCCACCCCACTCGGCCATCCATC	PstI

<b>ORF122</b>	Forward Reverse	AAAAAAGTCGAC-ATGCTTACCGCGCAAGCAGTTCCTC AAACTGCAG-TCAGGAACACAAACGATGACGAATATCCGTATC	Sall PstI
<b>ORF125</b>	Forward Reverse	AAAGAATTC-GCGCTGTTTTTTGCGGCGGCGTAT AAACTGCAG-CGCCGTTTCAAGACGAAAAAGTCG	EcoRI PstI
<b>ORF126</b>	Forward Reverse	AAAGAATTC-GCGGAAACGGTCAAG AAACTGCAG-TTAATCTTGCTTCCGATATAC	EcoRI PstI
<b>ORF127</b>	Forward Reverse	AAAGAATTC-ATGACTGATAATCGGGGGTTTACG AAAAAAGTCGAC-CTTAAGTAACTTGCAGTCCTTATC	EcoRI Sall
<b>ORF128</b>	Forward Reverse	AAAGAATTC-ATGCAAGCTGTCCGCTACAGGCC AAACTGCAG-CTATTGCAATGCGCCGCCGCGGAATGTTGAGCAGGCG	EcoRI PstI
<b>ORF129</b>	Forward Reverse	AAAGAATTC-ATGGATTTTCGTTTTGACATTATTTACGAATACCG AAACTGCAG-TTATTTTTTGATGAAATTTTGGGGCGG	EcoRI PstI
<b>ORF130</b>	Forward Reverse	AAAGAATTC-GCAGTACTTGCCAT TCTCGGTGCG AAACTGCAG-CTCCGGATCGTCTGTAAACGCATT	EcoRI PstI
<b>ORF 131</b>	Forward Reverse	GCGGATCCCATATG-GAAATTCGGGCAATAAAAT CCCGCTCGAG-CCAGCGGACGCGTTC	BamHI-NdeI XhoI
<b>ORF 132</b>	Forward Reverse	GCGGATCCCATATG-AAAGAAGCGGGGTTTG CCCGCTCGAG-CCAATCTGCCAGCCGT	BamHI-NdeI XhoI
<b>ORF 133</b>	Forward Reverse	CGCGGATCCCATATG-GAAGATGCAGGGCGCG CCCGCTCGAG-AAACTGTAGCTCATCGT	BamHI-NdeI XhoI
<b>ORF 134</b>	Forward Reverse	GCGGATCCCATATG-TCTGTGCAAGCAGTATTG CCCGCTCGAG-ATCCTGTGCCAATGCG	BamHI-NdeI XhoI
<b>ORF 135</b>	Forward Reverse	GCGGATCCCATATG-CCGTCTGAAAAAGCTTT CCCGCTCGAG-AAATACCGCTGAGGATG	BamHI-NdeI XhoI
<b>ORF 136</b>	Forward Reverse	CGCGGATCCGCTAGC-ATGAAGCGGCGTATAGCC CCCGCTCGAG-TTCCGAATATTTGGAACCTTT	BamHI-NheI XhoI
<b>ORF 137</b>	Forward Reverse	CGCGGATCCCATATG-GGCACGGCGGGAAATA CCCGCTCGAG-ATAACGGTATGCCGCC	BamHI-NdeI XhoI
<b>ORF 138</b>	Forward Reverse	GCGGATCCCATATG-TTTCGTTTACAATTCAGGC CCCGCTCGAG-CGGCGTTTTATAGCGG	BamHI-NdeI XhoI
<b>ORF 139</b>	Forward Reverse	GCGGATCCCATATG-GCTTTTTTGGCGGTAATG CCCGCTCGAG-TAACGTTTCCGTGCGTTT	BamHI-NdeI XhoI



<b>ORF 140</b>	Forward Reverse	GCGGATCCCATATG-TTGCCACAGGCAGC CCCGCTCGAG-GACGATGGCAAACAGC	BamHI-NdeI XhoI
<b>ORF 141</b>	Forward Reverse	GCGGATCCCATATG-CCGTCTGAAGCAGTCT CCCGCTCGAG-ATCTGTTGTTTTTAAATATT	BamHI-NdeI XhoI
<b>ORF 142</b>	Forward Reverse	GCGGATCCCATATG-GATAATTCTGGTAGTGAAG CCCGCTCGAG-AAACGTATAGCCTACCT	BamHI-NdeI XhoI
<b>ORF 143</b>	Forward Reverse	GCGGATCCCATATG-GATACCGCTTTGAACCT CCCGCTCGAG-AATGGCTTCCGCAATATG	BamHI-NdeI XhoI
<b>ORF 144</b>	Forward Reverse	GCGGATCCCATATG-ACCTTTTTACAACGTTTGC CCCGCTCGAG-AGATTGTTGTTGTTTTTCG	BamHI-NdeI XhoI
<b>ORF 147</b>	Forward Reverse	GCGGATCCCATATG-TCTGTCTTTCAAACGGC CCCGCTCGAG-TTTGTTTTTGCAAGACAG	BamHI-NdeI XhoI

NB:

- restriction sites are underlined
- for ORFs 110-130, where the ORF itself carries an *EcoRI* site (eg. ORF122), a *SalI* site was used in the forward primer instead. Similarly, where the ORF carries a *PstI* site (eg. ORFs 115 and 127), a *SalI* site was used in the reverse primer.

TABLE II – Summary of cloning, expression and purification

ORF	PCR/cloning	His-fusion expression	GST-fusion expression	Purification
orf 1	+	+	+	His-fusion
orf 2	+	+	+	GST-fusion
orf 2.1	+	n.d.	+	GST-fusion
orf 4	+	+	+	His-fusion
orf 5	+	n.d.	+	GST-fusion
orf 6	+	+	+	GST-fusion
orf 7	+	+	+	GST-fusion
orf 8	+	n.d.	n.d.	
orf 9	+	+	+	GST-fusion
orf 10	+	n.d.	n.d.	
orf 11	+	n.d.	n.d.	
orf 13	+	n.d.	+	GST-fusion
orf 15	+	+	+	GST-fusion
orf 17	+	n.d.	n.d.	
orf 18	+	n.d.	n.d.	
orf 19	+	n.d.	n.d.	
orf 20	+	n.d.	n.d.	
orf 22	+	+	+	GST-fusion
orf 23	+	+	+	His-fusion
orf 24	+	n.d.	n.d.	
orf 25	+	+	+	His-fusion
orf 26	+	n.d.	n.d.	
orf 27	+	+	+	GST-fusion
orf 28	+	+	+	GST-fusion
orf 29	+	n.d.	n.d.	
orf 32	+	+	+	His-fusion
orf 33	+	n.d.	n.d.	
orf 35	+	n.d.	n.d.	
orf 37	+	+	+	GST-fusion
orf 58	+	n.d.	n.d.	
orf 65	+	n.d.	n.d.	
orf 66	+	n.d.	n.d.	
orf 72	+	+	n.d.	His-fusion
orf 73	+	n.d.	+	n.d.
orf 75	+	n.d.	n.d.	
orf 76	+	+	n.d.	His-fusion
orf 79	+	+	n.d.	His-fusion
orf 83	+	n.d.	+	n.d.
orf 84	+	n.d.	n.d.	

orf 85	+	n.d.	+	GST-fusion
orf 89	+	n.d.	+	GST-fusion
orf 97	+	+	+	GST-fusion
orf 98	+	n.d.	n.d.	
orf 100	+	n.d.	n.d.	
orf 101	+	n.d.	n.d.	
orf 102	+	n.d.	n.d.	
orf 103	+	n.d.	n.d.	
orf 104	+	n.d.	n.d.	
orf 105	+	n.d.	n.d.	
orf 106	+	+	+	His-fusion
orf 109	+	n.d.	n.d.	
orf 110	+	n.d.	n.d.	
orf 111	+	+	n.d.	His-fusion
orf 113	+	+	n.d.	His-fusion
orf 115	n.d.	n.d.	n.d.	
orf 119	+	+	n.d.	His-fusion
orf 120	+	+	n.d.	His-fusion
orf 121	+	n.d.	n.d.	
orf 122	+	+	n.d.	His-fusion
orf 125	+	+	n.d.	His-fusion
orf 126	+	+	n.d.	His-fusion
orf 127	+	+	n.d.	His-fusion
orf 128	+	n.d.	n.d.	
orf 129	+	+	n.d.	His-fusion
orf 130	+	n.d.	n.d.	
orf 131	+	+	+	n.d.
orf 132	+	+	+	His-fusion
orf 133	+	n.d.	+	GST-fusion
orf 134	+	n.d.	n.d.	
orf 135	+	n.d.	n.d.	
orf 136	+	n.d.	n.d.	
orf 137	+	n.d.	+	GST-fusion
orf 138	+	n.d.	+	GST-fusion
orf 139	+	n.d.	n.d.	
orf 140	+	n.d.	n.d.	
orf 141	+	n.d.	n.d.	
orf 142	+	n.d.	n.d.	
orf 143	+	n.d.	n.d.	
orf 144	+	n.d.	+	n.d.
orf 147	+	n.d.	n.d.	

**CLAIMS**

1. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, and 8.
2. A nucleic acid molecule which encodes a protein according to claim 1.
- 5 3. A nucleic acid molecule according to claim 2, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, and 7.
4. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 10 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 15 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 20 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 25 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892..
- 30 5. A protein having 50% or greater sequence identity to a protein according to claim 4.

6. A protein comprising a fragment of an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892..
7. An antibody which binds to a protein according to any one of claims 4 to 6.
8. A nucleic acid molecule which encodes a protein according to any one of claims 4 to 6.
9. A nucleic acid molecule according to claim 8, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289,

291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329,  
331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369,  
371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409,  
411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449,  
5 451, 453, 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489,  
491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529,  
531, 533, 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569,  
571, 573, 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609,  
611, 613, 615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649,  
10 651, 653, 655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689,  
691, 693, 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729,  
731, 733, 735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769,  
771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809,  
811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849,  
15 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889,  
& 891..

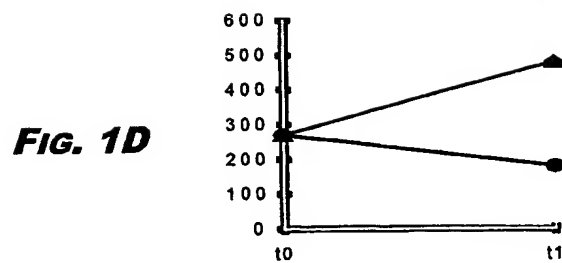
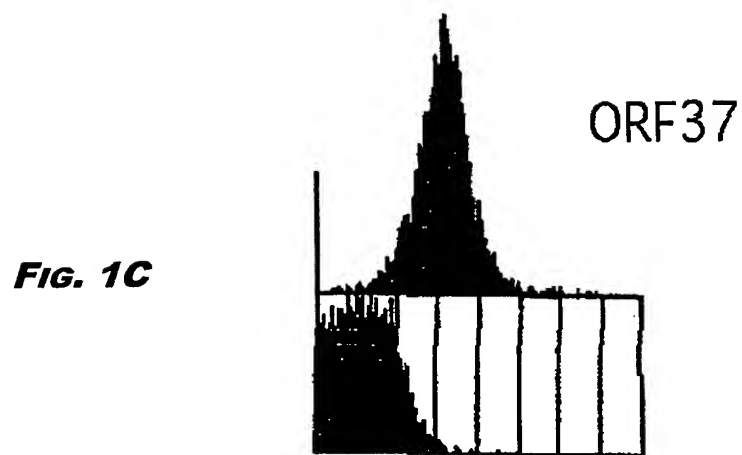
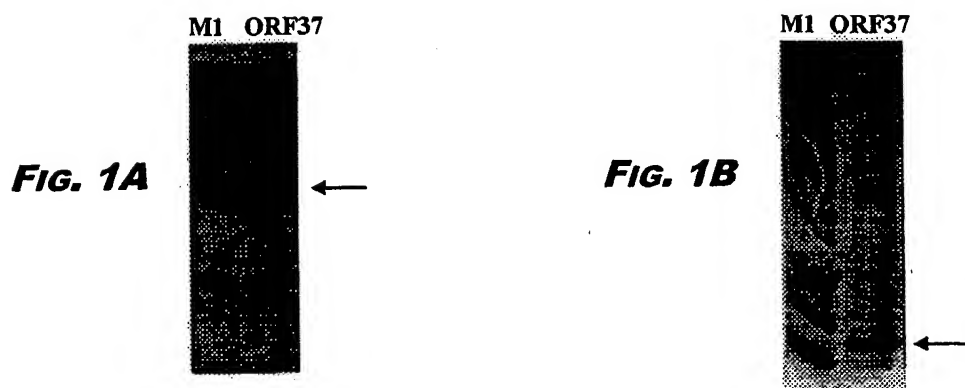
10. A nucleic acid molecule comprising a fragment of a nucleotide sequence selected from the  
group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39,  
41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91,  
20 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133,  
135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173,  
175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213,  
215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253,  
255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293,  
25 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333,  
335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373,  
375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413,  
415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453,  
455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 491, 493,  
30 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529, 531, 533,  
535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569, 571, 573,  
575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609, 611, 613,  
615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649, 651, 653,  
655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689, 691, 693,  
35 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729, 731, 733,

735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769, 771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809, 811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849, 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889, & 891..

- 5 11. A nucleic acid molecule comprising a nucleotide sequence complementary to a nucleic acid molecule according to any one of claims 8 to 10.
12. A nucleic acid molecule comprising a nucleotide sequences having 50% or greater sequence identity to a nucleic acid molecule according to any one of claims 8-11.
13. A nucleic acid molecule which can hybridise to a nucleic acid molecule according to any  
10 one of claims 8-12 under high stringency conditions.
14. A composition comprising a protein, a nucleic acid molecule, or an antibody according to any preceding claim.
15. A composition according to claim 14 being a vaccine composition or a diagnostic composition.
- 15 16. A composition according to claim 14 or claim 15 for use as a pharmaceutical.
17. The use of a composition according to claim 14 in the manufacture of a medicament for the treatment or prevention of infection due to Neisserial bacteria.

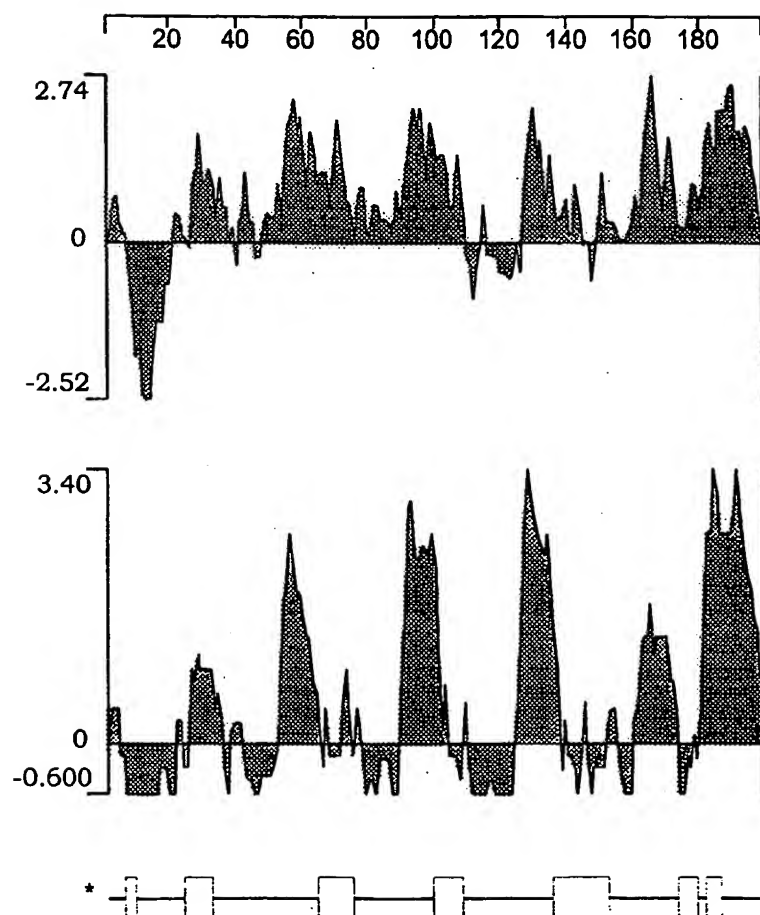
1/24

**FIGURE 1**





2/24

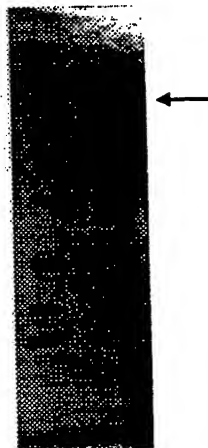
**FIG 1E**

3/24

**FIGURE 2**

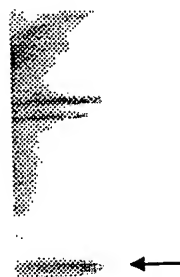
**FIG. 2A**

M1 ORF5



**FIG. 2B**

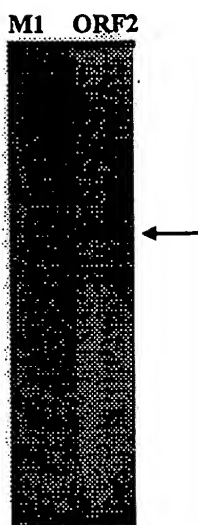
TP



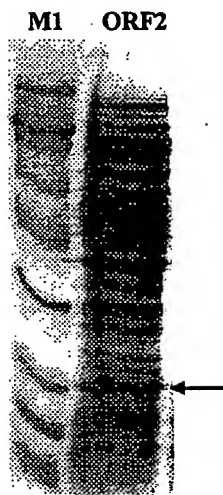
4/24

**FIGURE 3**

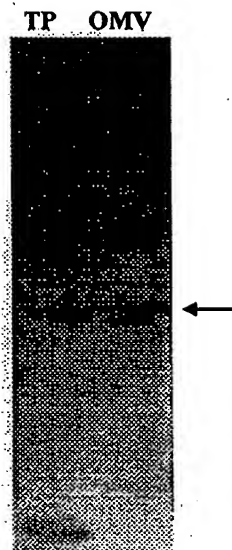
**FIG. 3A**



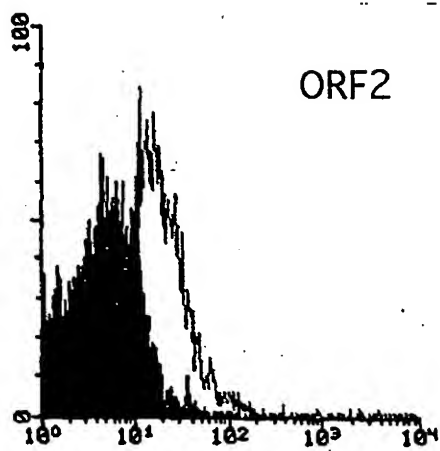
**FIG. 3B**



**FIG. 3C**



**FIG. 3D**

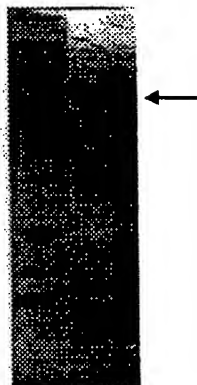


5/24

**FIGURE 4**

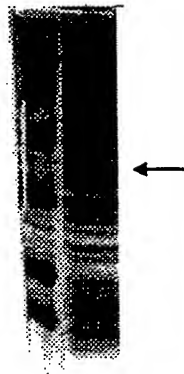
***Fig. 4A***

M1 ORF15



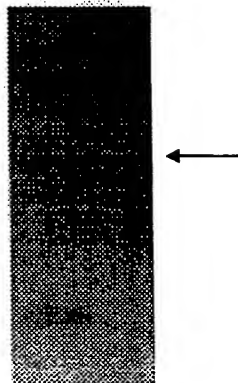
***Fig. 4B***

M2 ORF15

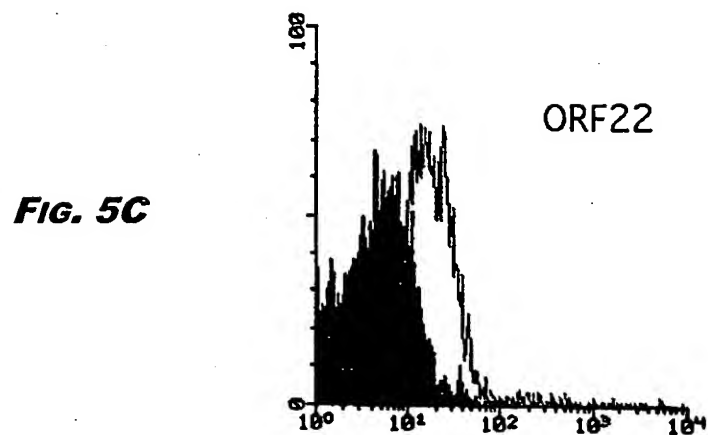
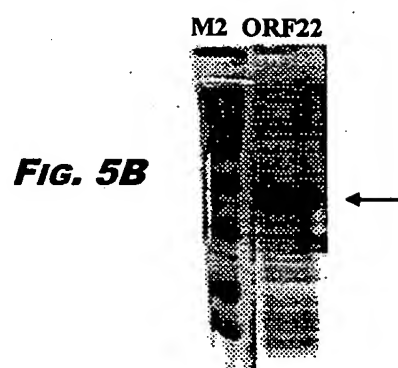
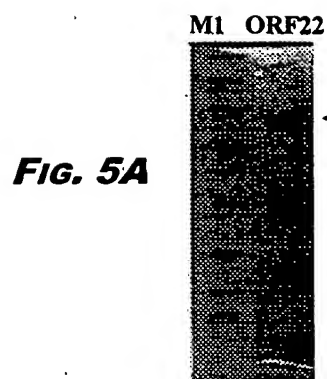


***Fig 4C***

TP OMV



6/24

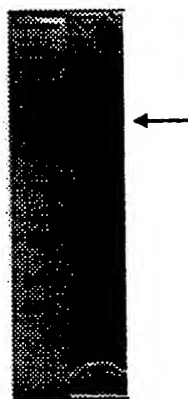
**FIGURE 5**

7/24

**FIGURE 6**

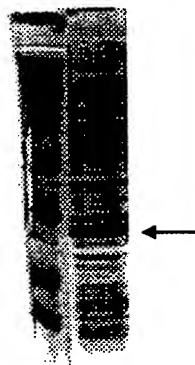
***Fig. 6A***

M1 ORF28



***Fig. 6B***

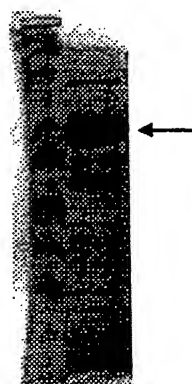
M2 ORF28



**FIGURE 7**

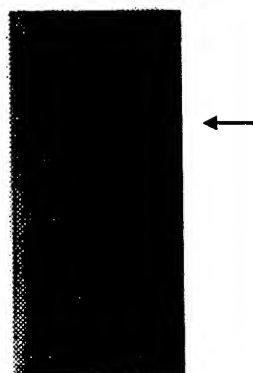
***Fig. 7A***

M1 ORF32

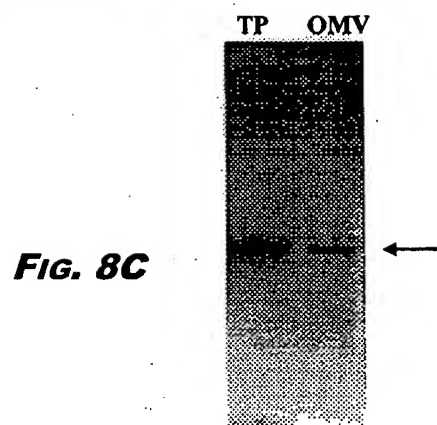
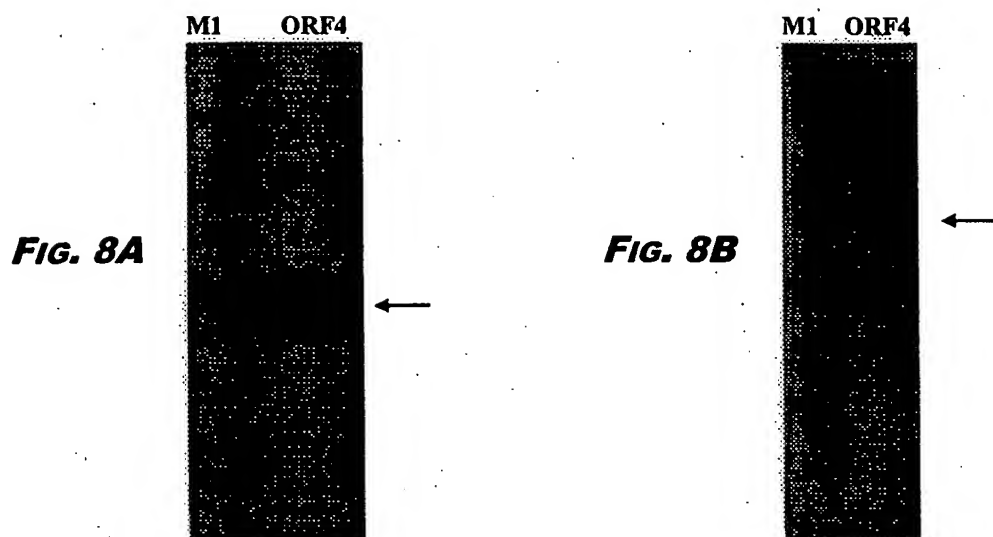
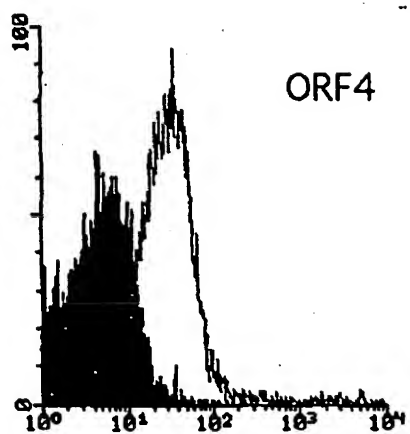


***Fig. 7B***

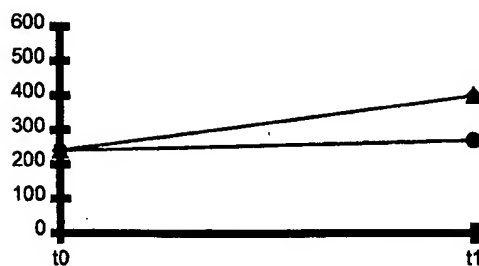
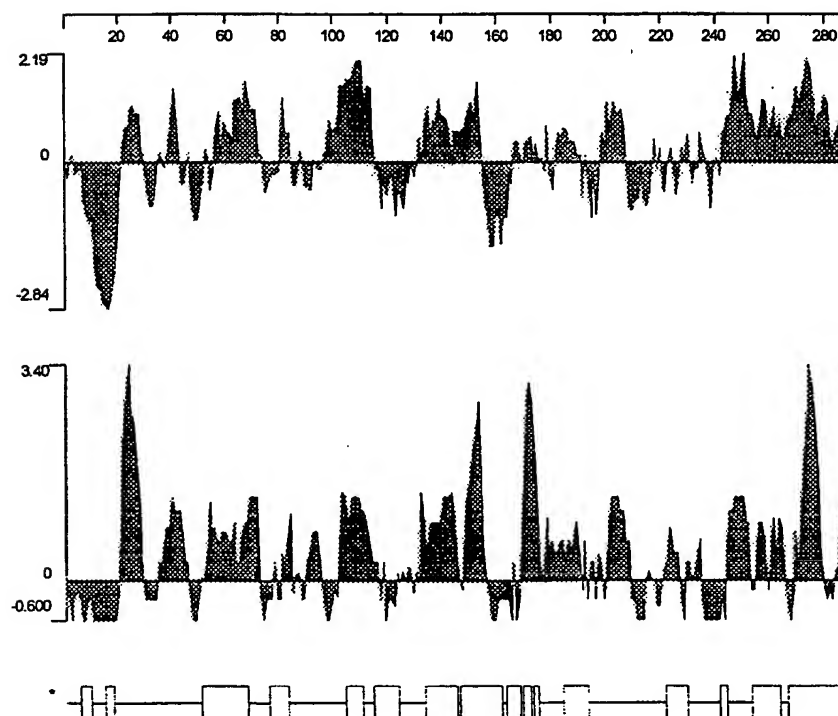
M1 ORF32



8/24

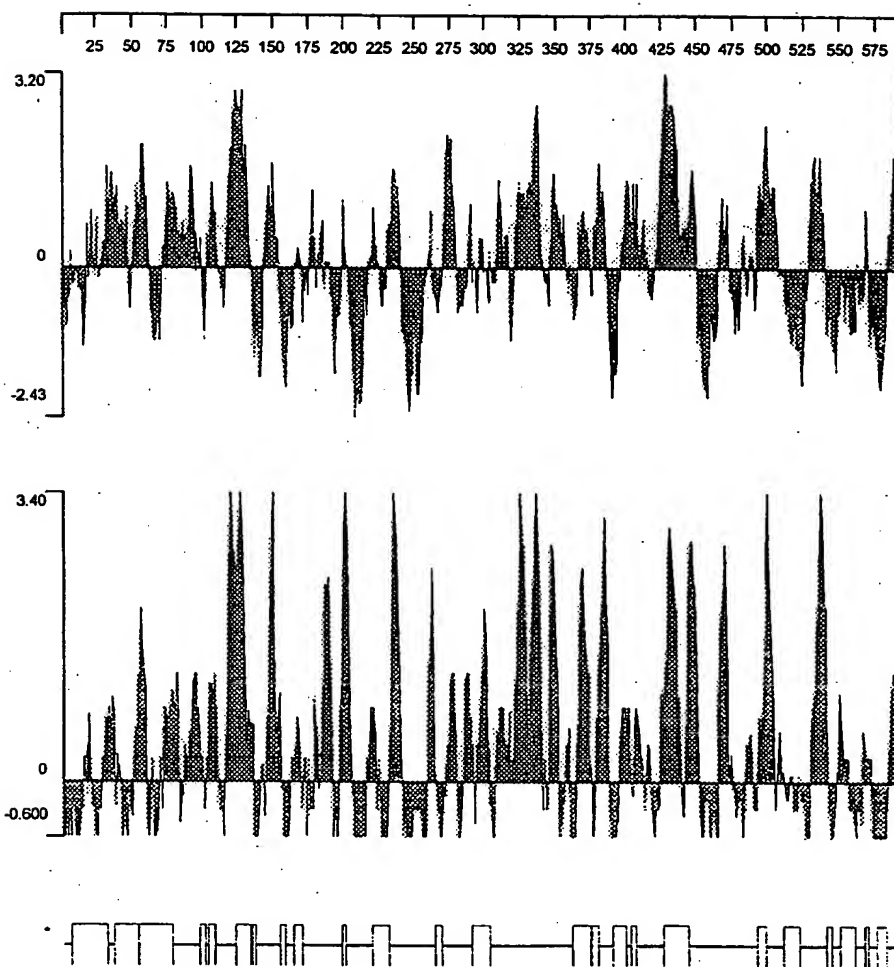
**FIGURE 8****FIG. 8D**

9/24

**Fig. 8E****Fig. 8F**



10/24

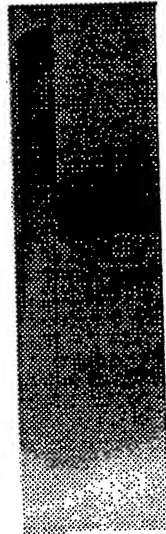
**FIGURE 9**

11/24

**FIGURE 10**

**Fig. 10A**

M1 ORF76

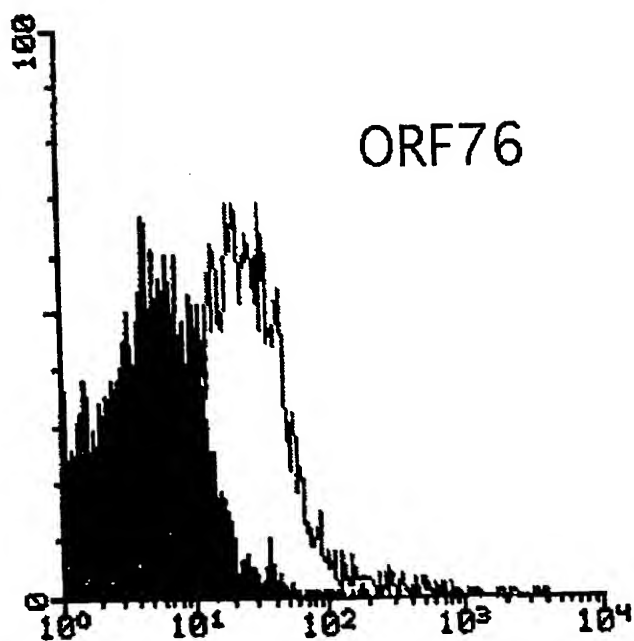


**Fig. 10B**

TP OMV



**Fig. 10C**



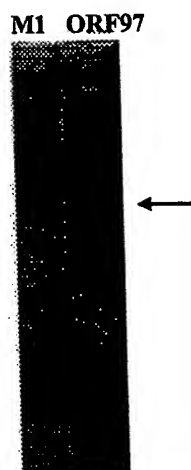
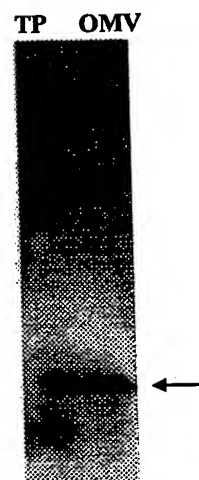
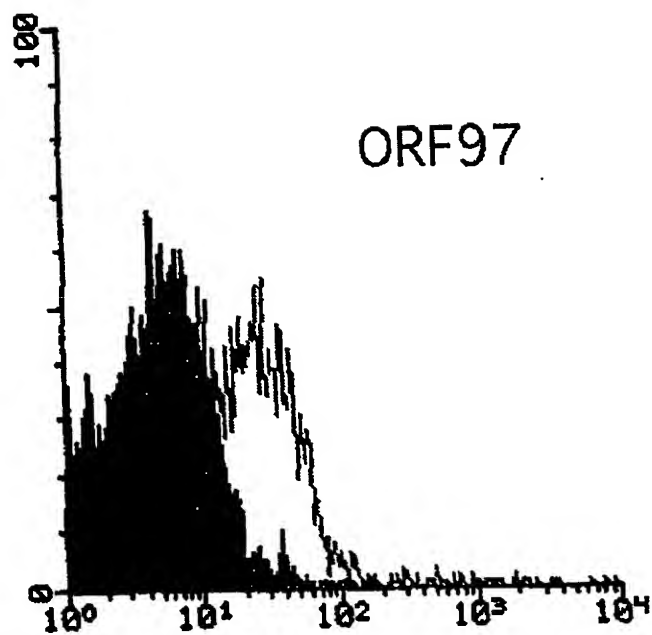
12/24

**FIGURE 11**

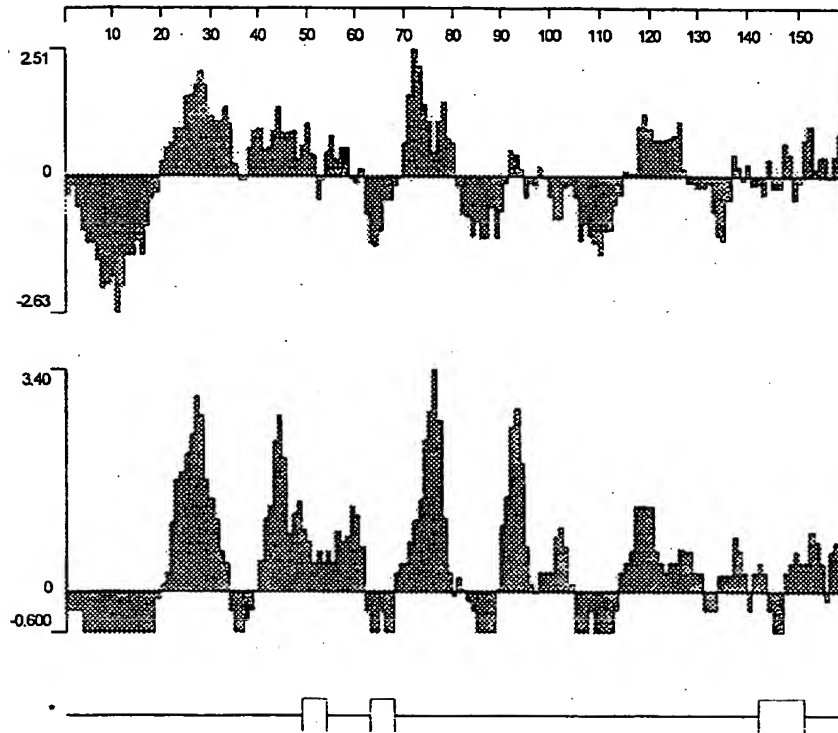
M1 ORF89



13/24

**FIGURE 12****FIG. 12A****FIG. 12B****FIG. 12C****FIG. 12D**

14/24

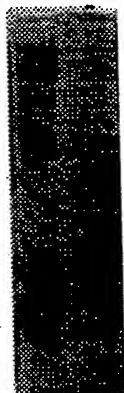
**FIG. 12E**

15/24

**FIGURE 13**

**Fig. 13A**

M1 ORF106

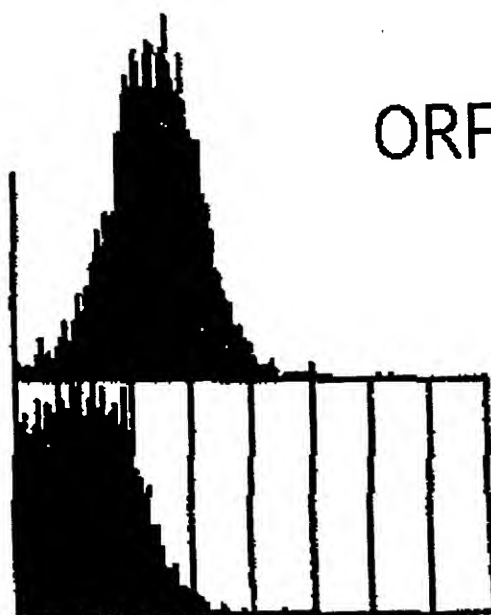


**Fig. 13B**

M2 ORF106

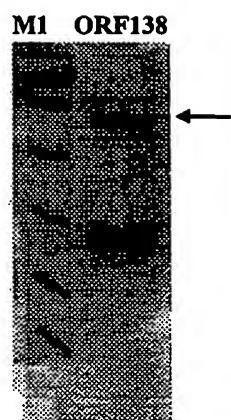
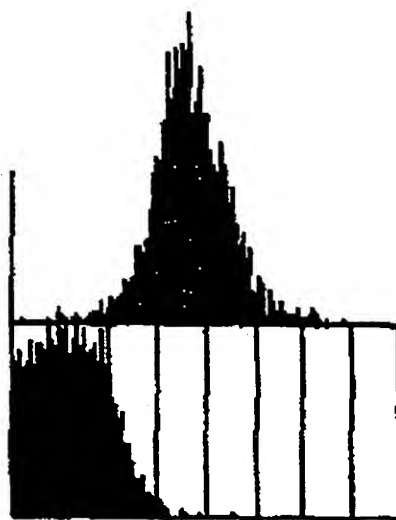


**Fig. 13C**



ORF 106

16/24

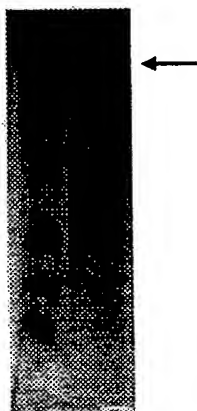
**FIGURE 14****FIG. 14A****FIG. 14B**

17/24

**FIGURE 15**

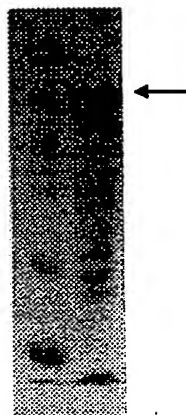
***Fig. 15A***

M1 ORF23



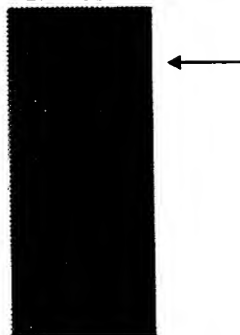
***Fig. 15B***

M2 ORF23



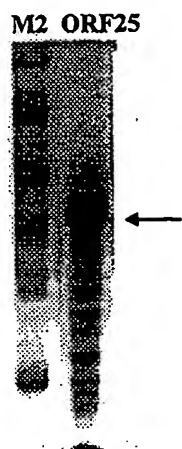
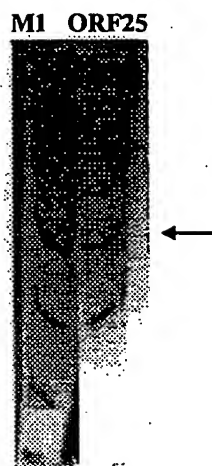
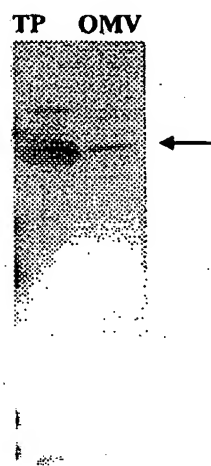
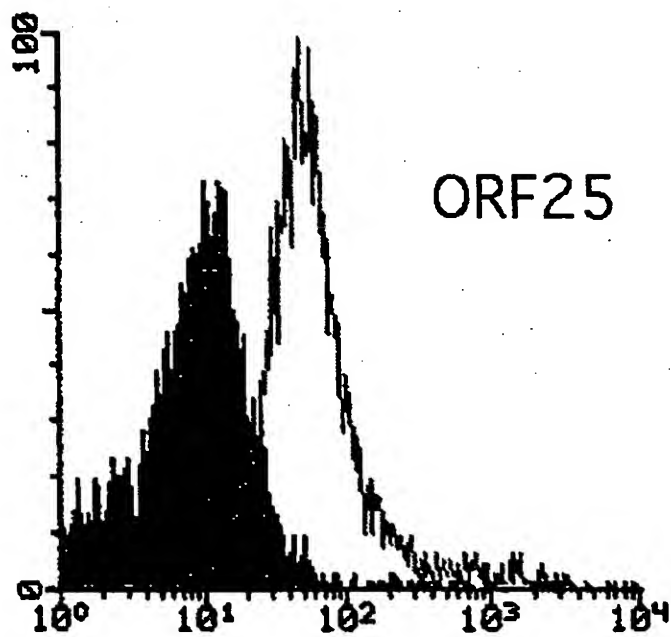
***Fig 15C***

TP OMV

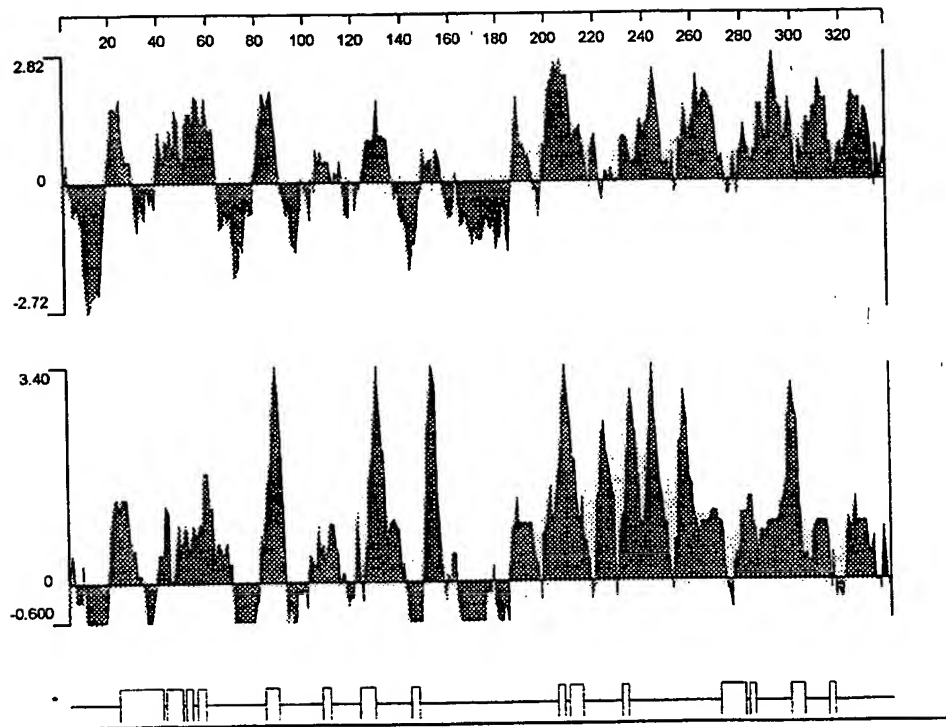




18/24

**FIGURE 16****FIG. 16A****FIG. 16B****FIG. 16C****FIG. 16D**

19/24

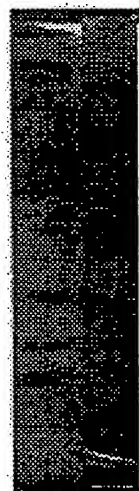
**FIG. 16E**

20/24

**FIGURE 17**

***FIG. 17A***

M1 ORF27



***FIG. 17B***

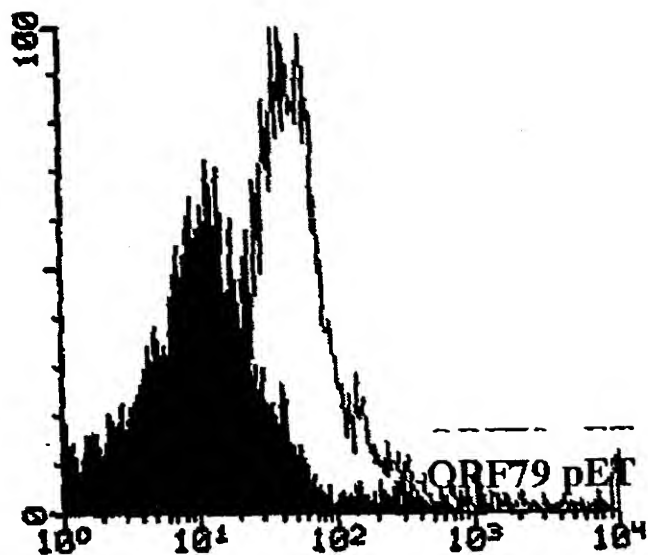
M2 ORF27



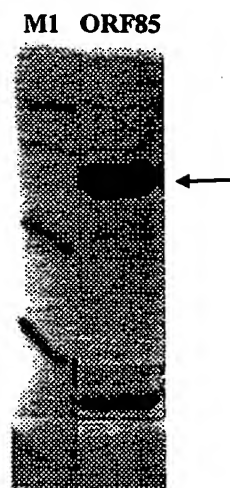
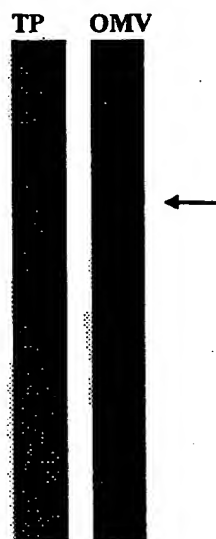
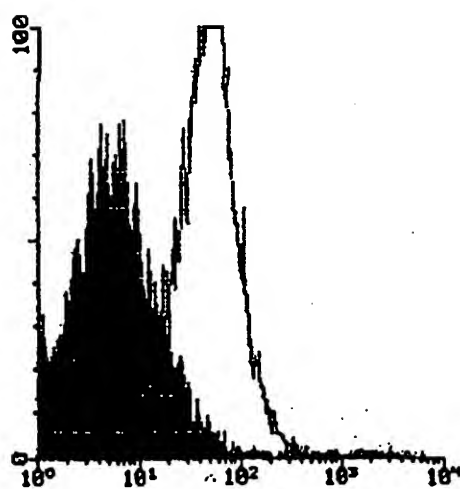
21/24

**FIGURE 18****Fig. 18A**

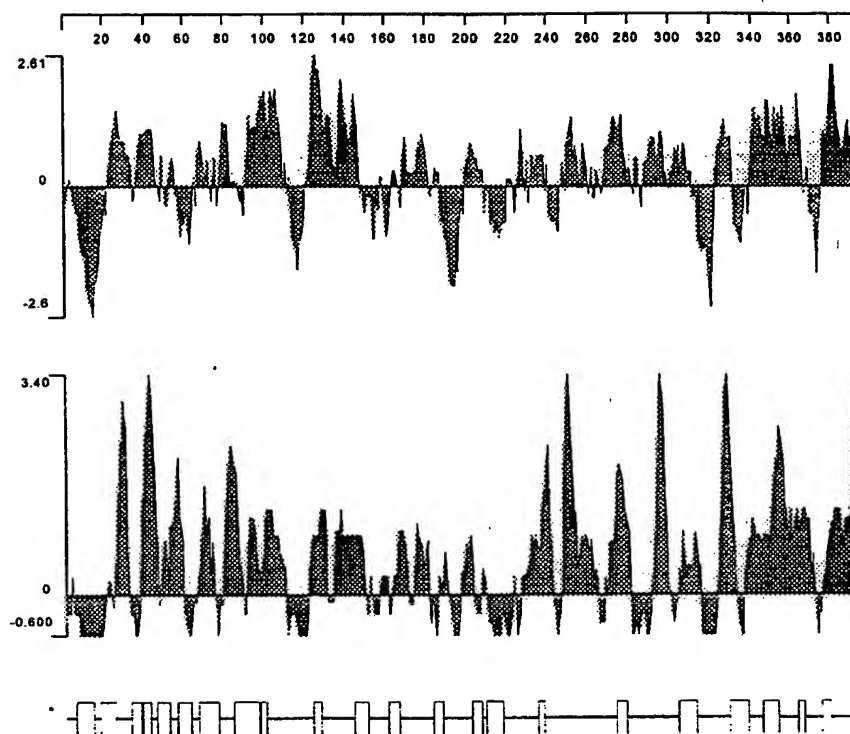
M1 ORF79

**Fig. 18B**

22/24

**FIGURE 19****Fig. 19A****Fig. 19B****FIG. 19C**

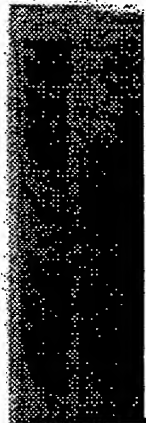
23/24

**Fig 19D**

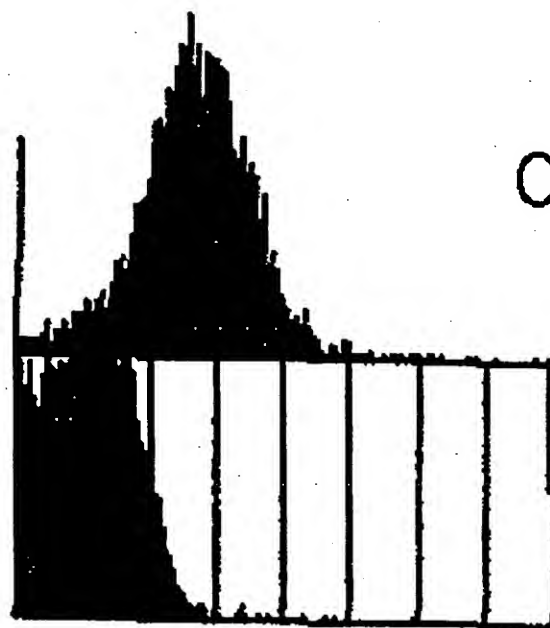
24/24

**FIGURE 20****Fig. 20A**

M1 ORF132

**Fig. 20B**

M2 ORF132

**Fig. 20C**

ORF132



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> C12N 15/31, C07K 14/22, 16/12, G01N 33/53, A61K 39/095 // (C12N 15/31, C12 R 1:36)	<b>A3</b>	<b>(11) International Publication Number:</b> <b>WO 99/24578</b>  <b>(43) International Publication Date:</b> 20 May 1999 (20.05.99)																					
<b>(21) International Application Number:</b> PCT/IB98/01665  <b>(22) International Filing Date:</b> 9 October 1998 (09.10.98)  <b>(30) Priority Data:</b> <table border="0"> <tr> <td>9723516.2</td> <td>6 November 1997 (06.11.97)</td> <td>GB</td> </tr> <tr> <td>9724190.5</td> <td>14 November 1997 (14.11.97)</td> <td>GB</td> </tr> <tr> <td>9724386.9</td> <td>18 November 1997 (18.11.97)</td> <td>GB</td> </tr> <tr> <td>9725158.1</td> <td>27 November 1997 (27.11.97)</td> <td>GB</td> </tr> <tr> <td>9726147.3</td> <td>10 December 1997 (10.12.97)</td> <td>GB</td> </tr> <tr> <td>9800759.4</td> <td>14 January 1998 (14.01.98)</td> <td>GB</td> </tr> <tr> <td>9819016.8</td> <td>1 September 1998 (01.09.98)</td> <td>GB</td> </tr> </table>  <b>(71) Applicant (for all designated States except US):</b> CHIRON S.P.A. [IT/IT]; Via Fiorentina, 1, I-53100 Siena (IT).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> MASIGNANI, Vega [IT/IT]; Via Pantaneto, 105, I-53100 Siena (IT). RAP-PUOLI, Rino [IT/IT]; Via delle Rocche, 1, Vagliagli, I-53019 Castelnuovo Berardenga (IT). PIZZA, Mariagrazia [IT/IT]; Strada di Montalbuccio, 160, I-53100 Siena (IT). SCARLATO, Vincenzo [IT/IT]; Via Firenze, 3/37, I-53134 Colle Val d'Elsa (IT). GRANDI, Guido [IT/IT]; 9° Strada, 4, I-20090 Segrate (IT).		9723516.2	6 November 1997 (06.11.97)	GB	9724190.5	14 November 1997 (14.11.97)	GB	9724386.9	18 November 1997 (18.11.97)	GB	9725158.1	27 November 1997 (27.11.97)	GB	9726147.3	10 December 1997 (10.12.97)	GB	9800759.4	14 January 1998 (14.01.98)	GB	9819016.8	1 September 1998 (01.09.98)	GB	<b>(74) Agent:</b> HALLYBONE, Huw, George; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).  <b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>With international search report.</i>  <b>(88) Date of publication of the international search report:</b> 2 March 2000 (02.03.00)
9723516.2	6 November 1997 (06.11.97)	GB																					
9724190.5	14 November 1997 (14.11.97)	GB																					
9724386.9	18 November 1997 (18.11.97)	GB																					
9725158.1	27 November 1997 (27.11.97)	GB																					
9726147.3	10 December 1997 (10.12.97)	GB																					
9800759.4	14 January 1998 (14.01.98)	GB																					
9819016.8	1 September 1998 (01.09.98)	GB																					
<b>(54) Title:</b> NEISSERIAL ANTIGENS  <b>(57) Abstract</b>  The invention provides proteins from <i>Neisseria meningitidis</i> (strains A and B) and from <i>Neisseria gonorrhoeae</i> including amino acid sequences, the corresponding nucleotide sequences, expression data, and serological data. The proteins are useful antigens for vaccines, immunogenic compositions, and/or diagnostics.																							



**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## INTERNATIONAL SEARCH REPORT

International Application No

PC1/IB 98/01665

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/31 C07K14/22 C07K16/12 G01N33/53 A61K39/095  
 //(C12N15/31,C12R1:36)

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C07K G01N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
E	<p>WO 99 55873 A (SMITHKLINE BEECHAM            BIOLOGICALS S.A.)            4 November 1999 (1999-11-04)            SEQ ID NOS: 1, 2, 3 and 4            page 82 -page 96            claims 3-5,10,11,17-21            ---            -/--</p>	<p>5-8,            10-16</p>

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

## \* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

1 December 1999

Date of mailing of the international search report

08 December 1999 (08.12.99)

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
 NL - 2280 HV Rijswijk  
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
 Fax: (+31-70) 340-3016

Authorized officer

Fuchs, U

## INTERNATIONAL SEARCH REPORT

International Application No

PCT, IB 98/01665

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>ROKBI, B. ET AL.: "Evaluation of Recombinant Transferrin-Binding Protein B Variants from Neisseria meningitidis for Their Ability To Induce Cross-Reactive and Bactericidal Antibodies against a Genetically Diverse Collection of Serogroup B Strains"</p> <p>INFECTION AND IMMUNITY, vol. 65, no. 1, January 1997 (1997-01), pages 55-63, XP002086937 abstract</p> <p>page 55, column 1, line 1 -page 56, column 2, line 78</p> <p>page 57, column 1, line 31 -page 61, column 2, line 63</p> <p>page 59; figure 3; table 2</p> <p>---</p>	1-17
A	<p>WO 96 12020 A (OREGON HEALTH SCIENCES UNIVERSITY) 25 April 1996 (1996-04-25) abstract</p> <p>page 32 -page 38; examples 2-4</p> <p>page 42 -page 44; example 9</p> <p>page 45; table 3</p> <p>---</p>	1-17
A	<p>WO 96 31618 A (THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL) 10 October 1996 (1996-10-10) abstract</p> <p>page 19, line 21 -page 20, line 13</p> <p>page 20, line 22 -page 21, line 11</p> <p>page 22, line 23 -page 24, line 16</p> <p>-----</p>	1-17

# INTERNATIONAL SEARCH REPORT

In. ational application No.  
PCT/IB 98/01665

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☒ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:  
  
((1-3) completely) and ((4-17) partially)
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☒ No protest accompanied the payment of additional search fees.

**FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210**

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

**Invention 1. Claims: ((1-3) completely) and ((4-17) partially)**

A protein comprising an amino acid sequence selected from the group consisting of SEQ ID NOS: 2,4,6 and 8 or fragments thereof; a protein having 50% or greater sequence identity to said protein; an antibody binding said protein; a nucleic acid encoding said protein; a nucleic acid comprising a sequence selected from the group consisting of SEQ ID NOS: 1,3,5 and 7 or fragments thereof; a composition comprising said protein, said nucleic acid or said antibody; the use of said composition;

**Invention 2 to 104. Claims (4-17) partially**

Idem as subject 1 but limited to the ORFs corresponding to examples 2-104 characterized by SEQ ID NOS: 9-892.

(Invention 2 is limited to SEQ ID NOS: 9-10; Invention 3 is limited to SEQ ID NOS: 11-18; Invention 4 is limited to SEQ ID NOS: 19-28; .....; Invention 104 is limited to SEQ ID NOS: 885-892).

In view of additional search fees paid, Inventions 5, 26, 55, 77 and 91 have been further searched.

# INTERNATIONAL SEARCH REPORT

I nformation on patent family members

International Application No

PC1, 1B 98/01665

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 9955873	A	04-11-1999	NONE	
WO 9612020	A	25-04-1996	US 5698438 A	16-12-1997
			AU 705509 B	27-05-1999
			AU 4007395 A	06-05-1996
			CA 2203116 A	25-04-1996
			EP 0793720 A	10-09-1997
			FI 971634 A	16-06-1997
			HU 77048 A	02-03-1998
			JP 10508469 T	25-08-1998
			NO 971768 A	03-06-1997
WO 9631618	A	10-10-1996	AU 5537096 A	23-10-1996
			CA 2217522 A	10-10-1996
			EP 0830456 A	25-03-1998
			JP 11503322 T	26-03-1999